The National Park Service
Alaska Region
Inventory & Monitoring Program

Data Management Plan
Central Alaska Network

National Park Service
201 1st Avenue
Fairbanks, AK 99701

December 2004
DRAFT

File Name: CAKN_DataMgmtPlan_12152004.doc

Recommended Citation:
Wilder, D.,editor, 2004. Data Management Plan, Central Alaska Network. National Park Service. Inventory and Monitoring Program, Fairbanks, AK. 86 pg, 12 appendices.

Topic(s):
Data management, Planning, Administrative, Interdisciplinary

Theme Keywords:
Reports, protocols, guidelines, plans, data

Placename Keywords:
Alaska, Central Alaska Network, Denali, Yukon-Charley Rivers, Wrangell-St. Elias

Acronyms:

AKRO	Alaska Regional Office
ARLIS	Alaska Resource Library and Information Services
DMP	Data Management Plan
GIS	Geographic Information System
I&M	Inventory & Monitoring (Program)
ITIS	Integrated Taxonomic Information System
NPS	National Park Service
PDF	Adobe Portable Document Format
CAKN	Central Alaska Network
YUGA	Yukon-Charley Rivers, Gates of the Arctic main office
LAN	Local Area Network
ERD	Entity Relationship Diagram (for databases)

Initial Distribution:
Website: *www.nature.nps.gov/im/units/cakn/DataMgt.htm*

Data Management Plan
Central Alaska Network

Alaska Region Approval Signatures

_____ _____
Sara Wesser, Regional Coordinator Date

Central Alaska Network Approval Signatures

_____ _____
Maggie MacCluskie, CAKN Coordinator Date

_____ _____
Doug Wilder, CAKN Data Manager Date

Table of Contents

List of Figures

List of Tables

Executive Summary

Chapter I: Introduction

- As a critical component of the CAKN Monitoring Program, the Data Management Plan (DMP) aims to:
    - Outline the long-term goals of a comprehensive data management strategy for the CAKN I&M Program
    - Associate data management goals with the long-term goals of the network and service-wide I&M program
    - Outline the procedures and work practices that support effective data management
    - Guide current and future staff of the CAKN to ensure that sound data management practices are followed
    - Guide the enhancement of legacy data to match formats and standards put forth in this plan
    - Encourage effective data management practices as an integral part of project management so all data are made available and usable for park management decisions now and into the future
    - Optimize and promote interagency sharing and development of data, software applications and analyses
    - Establish roles and responsibilities of program staff for managing data
    - Identify necessary elements for a functional data management program and describe any anticipated changes to those elements
    - Establish an organization schema for Program data and information so that they are retrievable by staff, cooperators and the public
    - Establish basic quality control standards
    - Establish standards for data, data distribution and data archiving to ensure the long-term integrity of data, associated metadata and any supporting information

- Specific data and information the CAKN Program deems necessary to meet objectives includes:
    - Core variable data measured in the field
    - Data derived via vital sign protocols from core variable data
    - Spatial data files
    - Photographs (field and aerial)
    - Laboratory data
    - “Data” or “Technical” Reports including protocols
    - Administrative Reports
    - Field data sheets, books
    - Selected external and legacy data and datasets

- The CAKN data management strategy draws from national-level I&M guidelines and formalizes them as policy at the network level. More detailed data management strategies are documented in standard operating procedures specific to a given data collection effort.

- As a plan, the DMP will likely undergo significant revision as CAKN operations are implemented. Informal review and revision will be on-going to meet the

changing needs of the CAKN I&M Program. The latest version of the plan is available on the CAKN website at *www.nature.nps.gov/im/units/cakn/DataMgt.htm*.

Chapter II: Process and Workflow

- Core data management operations are conducted within the basic process and workflow of a data-generating project. Projects are divided into the following stages:
    1. Project Initiation
    2. Planning and Approval
    3. Design and Testing
    4. Implementation
    5. Product Delivery
    6. Product Integration
    7. Close Out and Evaluation

- The CAKN will use a Project Tracking database to track basic project parameters.

Chapter III: Management Roles and Responsibilities

- For the CAKN Program to work effectively, everyone within the Network will have stewardship responsibilities in the production, analysis, management, and/or end use of the data.

- Each Vital Sign monitoring protocol and any project study plan contains specific instructions for assignments and tasks that nest within the overall framework of the Data Management Plan. Individuals who carry out monitoring protocols and inventory study plans are responsible for reading and understanding these instructional guidelines.

- The fundamental role of the Network data manager is to understand and determine program and project requirements, to create and maintain data management infrastructure and standards, and to communicate and work with all responsible individuals.

- The Project Leader (PL) oversees and directs operations, including data management requirements, for one or more Network projects. The PL maintains communication with project staff, Network Data Manager, and resource specialist regarding data management. Project leaders are responsible for designating an alternate leader who is capable of maintaining project operations in his or her absence.

- The Network Coordinator interfaces with project leaders to ensure that timelines for data entry, validation, verification, summarization/analysis and reporting are met. Additionally, the Network Coordinator must review and approve proposed changes to project protocols prior to implementation.

Chapter IV: Infrastructure

- The CAKN monitoring program relies heavily on park, regional and national information technology (IT) personnel and resources to maintain the computer resource infrastructure.

- The CAKN data management "system" is currently (12/15/2004) in development. The system design presented here represents current plans which may be significantly altered by the point of full implementation.

- The CAKN will establish and operate a server database (MS SQL Server) housing electronic data and information managed by the program. This relational database will allow staff to browse, evaluate, export, analyze and integrate vital sign monitoring data and information for research, management and reporting purposes. Development of this system is planned in three stages culminating in conversion of all data to GeoDatabase format for serving via the ESRI Spatial Database Engine. GIS and web mapping applications will be implemented to enhance data access.

- The Alaska Regional Office provides the following that the CAKN will utilize to meet its goals:
    - The Wide Area Network file server for general file exchange and storage
    - GIS and related tabular data accessible via custom applications distributed to the parks as well as the Alaska GIS Data Clearing House (*www.nps.gov/akso/gis*).
    - File server to provide offsite storage for all CAKN data

Chapter V: Acquisition and Processing

- Data designs for all CAKN monitoring data and CAKN-initiated projects will conform to the I&M Natural Resource Template Standards and confine, where possible, data values to specific ranges.

- Data discovery is an on-going process requiring regular data searches and visits to Network parks in order to ensure that the CAKN I&M Program maintains as much relevant material pertaining to the parks as possible. Data acquired by the CAKN will be developed, if necessary, as digital datasets conforming to NPS and I&M database standards.

- Upon the completion of any field work, a summary of what was collected will be entered in a project tracking database.

- Acquisition of data sourced outside the NPS will be addressed in the appropriate Vital Sign protocol.

- Data and information for the CAKN may be organized into five categories: Monitoring, Inventory, Prototype (DENA LTEM), Short-term (e.g. pilot projects) and Legacy (primarily "mined" data).

- The CAKN intends to scan hardcopy references and materials, saving them as .pdf files, in order to create a digital library.

- Each CAKN project uses a working database developed in MS Access to perform initial data processing (entry, verification, validation, metadata generation) after acquisition.

- As part of data collection and entry duties, the project leader and data manager will ensure that data from laboratories match project data designs and formats. They will also build into the project data design a mechanism for relating instances of field work (data collection events), raw data, and laboratory results. This will be done via sample and specimen labels that can be matched with field work event identification numbers.

- Significant changes to the protocols dictating data acquisition methods must be approved by the project leader, network coordinator and the data manager. The network coordinator must evaluate the proposed changes and determine if additional peer review is required before approving.

Chapter VI: Assurance (QA) and Quality Control (QC)

- Data management for the CAKN Inventory and Monitoring program must ensure that our projects produce and maintain data of the highest possible quality. The CAKN will develop a comprehensive set of SOPs and tools for quality assurance and control in field procedures, data entry/validation/verification and data use (browsing, sub setting, downloading, analysis, etc.).

- NPS Director's Order #11B specifies standards that apply not only to NPS-generated information, but also to information provided by other parties to the NPS if the NPS disseminates or relies upon this information.

- Project subjects and goals will drive data quality needs and control the kinds of analysis and summarization that may be defensibly applied.

- Laboratories that will be entering analysis results for a given vital sign will be supplied with a copy of the database application so that data may be entered in the manner and format matching that of the rest of the data for a given monitoring parameter. While most professional laboratories exercise their own QA/QC procedures, results received by a project leader are subject to the same QA/QC measures exacted on other project data.

- As a standard part of database design, the CAKN will build into database tables, fields that track at the record-level who entered the data, precise entry time and the protocol version under which the data were collected.

- Data Collection: At a minimum, data will be collected on formatted, project-specific data sheets that reflect the overall data design for the project and maximize limitations on values that may be recorded for different parameters. Sheets will be designed to minimize the amount of writing necessary to effectively record observations.

- Data Entry: Data will be entered as soon as reasonably possible after collection by someone familiar with data collection. Data will be entered into pre-designed database applications that resemble field sheets and maximize error control. Data will not be entered into spread sheets. The CAKN will maximize the use of database programming to control data entry. To the extent possible, data entry will be automated.

- Data Verification: Data verification is carried out by staff sanctioned by the project leader who are ideally familiar with data collection and entry. One hundred percent of records will be verified against original source data. Ten percent of records will be reviewed after initial verification by the project manager. If errors are found, the entire data set verified again. A record of each dataset's verification process including number of verification iterations and results will be prepared by the project leader as part of formal metadata generation.

- Data Validation: Corrections or deletions as a result of data validation require notations in the original paper field records and in any copies made for data entry about how and why the data were changed. The CAKN will maximize the use of automated routines and/or data summary/visualization such as histograms, line plots, and basic statistics to reveal possible logic and range errors.

- The data manager will conduct periodic "spot checks" of random CAKN monitoring projects to ensure compliance with data management plan and project protocol QA/QC procedures.

Chapter VII: Documentation

- Formal, standard metadata that complies with federal and NPS standards serve as the principal documentation for CAKN data.

- The CAKN data manager will establish a standard operating procedure for metadata generation and maintenance to be used by all CAKN projects.

- In general, a single metadata document will apply to both raw and certified versions of the data. Metadata records will be stored with both hard copy and digital archive data.

- Generally, metadata will be created by the project leader, with assistance from the data manager.

- Metadata will be served to the internet via the CAKN primary data server out of Fairbanks as well as the national server in Ft. Collins (NR-GIS metadata server). The CAKN data manager ensures metadata are up-to-date on all servers.

Chapter VIII: Analysis and Reporting

- The CAKN strategy towards data analysis and reporting rests upon providing sufficient funding for these activities so that they occur promptly—that is, to report on the previous phenological year (Oct-Sept) by the following March.

- The CAKN will also focus on producing an annual integrated "State of the Parks" report that effectively communicates the changes and trends observed in each Vital Sign to our primary audience—the natural resource managers of each park.

- Each CAKN project protocol will include a standard operating procedure addressing data analysis. Development of these will:
    1. Use straightforward, minimally structured sampling designs.
    2. Work closely with statisticians in developing and implementing change detection analyses.

3. Provide adequate support to project leaders for data analysis (including hiring of staff).

- In general, analysis steps are:
  1. Summarization
  2. Outlier detection
  3. Change over time
  4. Vital Sign relationships
  5. Time series analysis

- The main audience for monitoring data is the resource managers of each network park and other managers in the National Park Service system, who will use the information to assist with their management decisions.

- The CAKN vision for reporting includes the following central themes: (1) We will prepare monitoring reports that are understandable and useful to our primary audience: park resource managers, (2) We will prepare reports promptly, and (3) All reports will be readily available.

Chapter IX: Dissemination

- Data management within the CAKN I&M Program aims to ensure that
  - Data are easily discoverable and obtainable
  - No data that have not been subjected to full quality control are released
  - Distributed data are accompanied by complete metadata which clearly establishes the data as a product of the NPS I&M Program
  - Sensitive data are identified and protected from unauthorized access and inappropriate use
  - A complete record of data distribution/dissemination is maintained

- All data and materials collected or generated using National Park Service personnel and funds become the property of the National Park Service.

- Any important findings from research and educational activities should be promptly submitted for publication. Authorship must accurately reflect the contributions of those involved.

- Investigators must share collections, data, results, and supporting materials with other researchers whenever possible. In exceptional cases, where collections or data are sensitive or fragile, access may be limited.

- Network staff will classify sensitive data on a case-by-case, project-by-project basis.

- All references to protected information are removed or obscured in any reports, publications, maps, or other public forum.

- Repositories for CAKN data, products and selected project items include:
  - CAKN Primary Data Server
  - NPS Natural Resources GIS Metadata and Data Store (NR-GIS)

  - NPS on-line bibliographic database (NatureBib)
  - NPS on-line document server (NPS Focus)
  - Alaska Resources Library and Information Services (ARLIS)
  - EPA STORET (on-line water quality database)
  - University of Alaska, Fairbanks Museum (primarily for plant specimen)

- In general, data will be available upon completion of analysis and reporting. Data for which analysis and reporting has not been completed but are otherwise certified (verified and validated) will be released no later than one year after certification.

Chapter X: Maintenance, Storage and Archiving

- CAKN data maintenance, storage and archiving procedures aim to ensure that data and related documents and materials (digital and physical) are
  - Kept up-to-date with regards to content and format such that the data are easily accessed and their heritage and quality easily learned.
  - Physically secure against environmental hazards, catastrophe, and human malice
  - Archived in a manner that expedites recovery if needed

- Most maintenance activity will involve active monitoring datasets however finalized CAKN-project data will also be maintained along with active data in a common relational database system.

- Primary digital data maintenance will be performed on the main CAKN server in Fairbanks.

- A catalogue of the data and information on the CAKN server will be maintained on the CAKN website and reflect changes or updates to datasets.

- All digital files on the primary CAKN server will be backed up daily in Fairbanks. These data will also be backed up via the regional network to a server in Anchorage at least weekly. All digital files will be restorable from backup sources in either Fairbanks or Anchorage. Both of these backup sources will be read-only and accessible for data restoration purposes only.

- Raw, certified and analyzed data (data products) will be archived and a common metadata file will be associated with each.

- Datasets that are considered complete and inactive will be saved on the primary CAKN server in both ASCII and native formats. Data are considered complete and inactive when accompanying metadata, as generated by the project leader or other authorized personnel, list the project status as “complete”.

- Hard copy project materials (including project reports) will be archived according to NPS standards, policy and procedure.

- Photographs from each CAKN project will be entered into a database where attributes such as electronic file name, keywords, project, description, photographer, date and location will be catalogued. Digital photo management

will generally follow guidelines established in a draft management strategy prepared by the Alaska Southeast and Southwest I&M Networks.

I. Introduction

Collecting natural resource data is our first step toward understanding the ecosystems within our National Parks. These ecosystems are evolving, as is our knowledge of them and how they work. We use these "raw" data to analyze, synthesize, and model aspects of ecosystems. In turn, we use our results and interpretations to make decisions about the Park's vital natural resources. Thus, *data* collected by researchers and maintained through sound data management practices will become *information* through analyses, syntheses, and modeling.

Information is the common currency among the many and various activities and people involved in stewardship projects throughout our National Park System. These projects include park planning, creating inventories, short-term and long-term monitoring, restoration, control of invasive species and other species management, fire management, trail and road maintenance, law enforcement, and the communication of natural resource information to the public.

The Inventory and Monitoring Program (I&M; *www.nature.nps.gov/im*) represents a long-term commitment by the National Park Service (NPS) to assess and document the status and trends of park ecological resources. In 1998, the National Parks Omnibus Management Act established a framework for the I&M Program which fully integrates natural resource monitoring and other scientific activities into the management processes of the National Park System.

The Omnibus Management Act (1998) charges the Secretary of the Interior to "continually improve the ability of the National Park Service to provide state-of-the-art management, protection, and interpretation of and research on the resources of the National Park System, "and to "… assure the full and proper utilization of the results of scientific studies for park management decisions." Section 5934 of the Act requires the Secretary of the Interior to develop a program of "inventory and monitoring of National Park System resources to establish baseline information and to provide information on the long-term trends in the condition of National Park System resources."

To carry out this mission, the NPS initiated a service-wide, natural resource Inventory and Monitoring Program encompassing approximately 270 park units with significant natural resources. These park units are grouped based on similar ecology into 32 networks. Each network works towards the following programmatic goals of the I&M Program:

- Establish natural resource inventory and monitoring as a standard practice throughout the National Park system that transcends traditional program, activity, and funding boundaries.
- Inventory the natural resources and park ecosystems under National Park Service stewardship to determine their nature and status.
- Monitor park ecosystems to better understand their dynamic nature and condition and to provide reference points for comparisons with other, altered environments.
- Integrate natural resource inventory and monitoring information into National Park Service planning, management, and decision making.

- Share National Park Service accomplishments and information with other natural resource organizations and form partnerships for attaining common goals and objectives.

The last two of these goals can only be achieved through the development of a modern information management infrastructure (e.g., staffing, hardware, software) and procedures to ensure that relevant natural resource data collected by NPS staff, cooperators, researchers and others are entered, quality-checked, analyzed, reported, archived, documented, cataloged, and made available to others for management decision-making, research, and education. This Data Management Plan (DMP) serves as the overarching strategy for achieving these goals. The plan supports I&M Program goals and objectives by ensuring that Program data are documented, secure, and remain accessible and useful indefinitely.

Central Alaska Network

The Central Alaska Network (CAKN) is composed of Wrangell-St. Elias National Park and Preserve, Denali National Park and Preserve, and Yukon-Charley Rivers National Preserve (Figure 1.1). The CAKN is working to build a holistic picture of change across the ecosystems of the network — specifically, to detect change in ecological components and in the relationships among those components. Beginning in 2000, the CAKN began a 5-year program of biological inventories, and in 2002, began planning a network-wide monitoring program. Each network is required to develop a formal Monitoring Plan prior to initiation of monitoring. A DMP is a required adjunct to each network's Monitoring Plan. This document has been prepared for submittal and review on the same schedule of the network's Monitoring Plan (MacCluskie and Oakley 2004).

Natural Resource Inventories

The natural resource inventory (second bullet, page 7) comprises twelve basic datasets (Appendix A) and forms the basis for planning and development of the Network's monitoring program. The biological inventories began in 2000, focus on freshwater fish, small mammals, and vascular plants. These biological inventories (*www.nature.nps.gov/im/units/cakn/Inventory.htm*) help meet programmatic I&M goals and provide the CAKN with ecological indices from which long-term monitoring may begin.

Data resulting from the natural resource inventories are part of the CAKN data universe (see Chapter V) and will be subjected to the same management standards as the monitoring data.

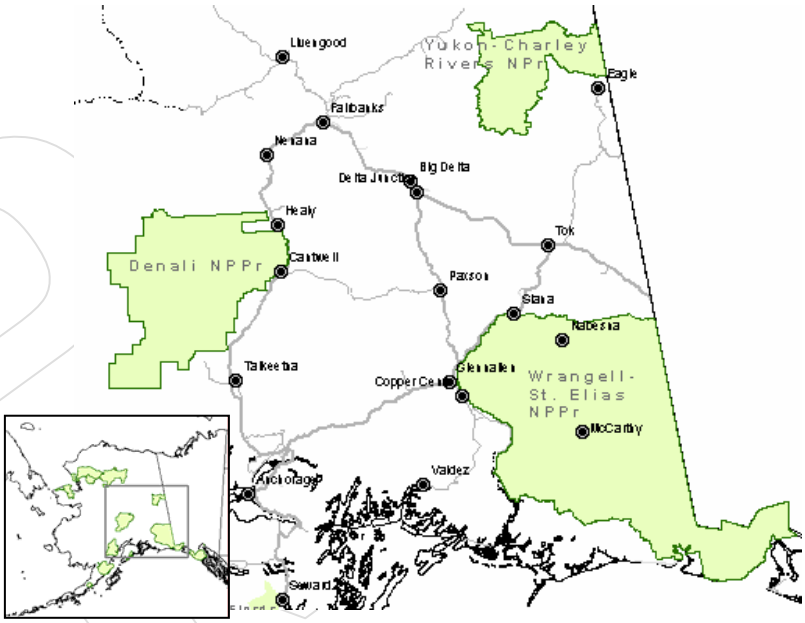


Figure 1.1. Park units of the Central Alaska Network.

I.1. Context of the Data Management Plan

A wealth of data and information management guidelines has been published or is generally available via the internet. The National Park Service's I&M Program guides data management by providing various guidelines (*www1.nrintra.nps.gov/im/datamgmt* – accessible via NPS computers only) to the 32 networks and parks in general. The CAKN data management strategy draws from this body of guidelines and formalizes them as policy at the network level. More detailed data management strategies are documented in standard operating procedures specific to a given data collection effort. These standard operating procedures fall under the aegis of this plan and adhere to the guidance, strategies and polices herein. Figure 1.2 depicts this data management guideline/policy inheritance model for the CAKN.

Detailed data management strategies for the CAKN will be drafted in light of this plan as part of the protocol development for a given data-gathering effort.

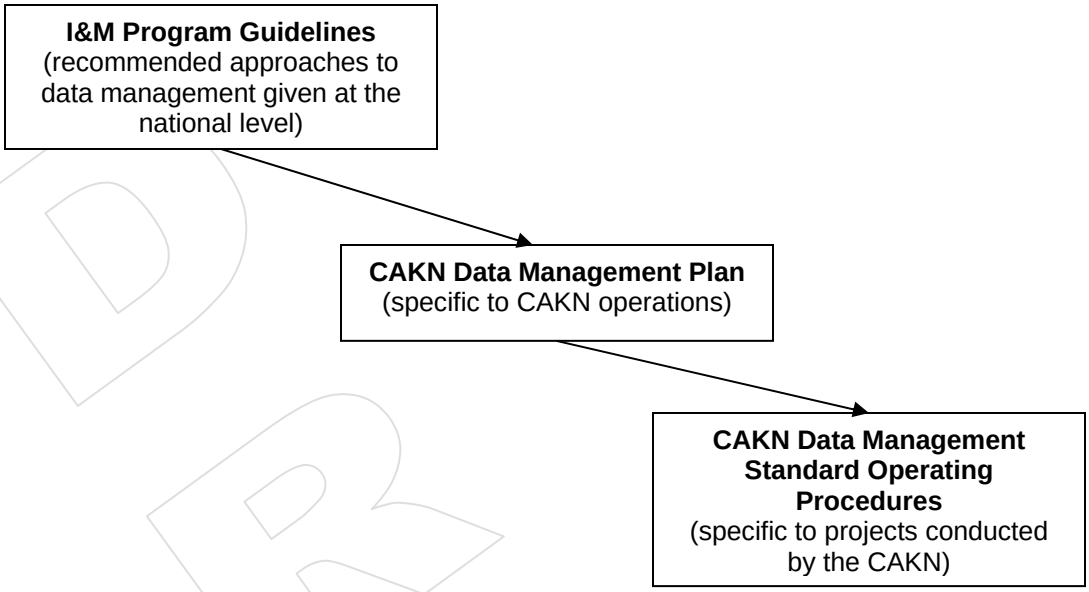


Figure 1.2. Context of national I&M Program guidelines, the CAKN Data Management Plan and detailed data management procedures for a project.

I.2. What are we managing?

Natural resource data are the vital building blocks for our evolving ecological understanding about park resources. But a set of data – whether collected the previous year or 20 years ago – must also be accompanied by sufficient context of how and why it was collected to maintain its value beyond the lifetimes of those who collected it. Therefore, a data management strategy cannot simply attend to the tables, fields, and values that make up a data set. There must also be a process for developing, preserving, and integrating the context that makes it interpretable and valuable.

The term "data" is frequently used in a way that also encompasses other products that are generated alongside the tabular and spatial data that are the primary targets of our data management efforts. These products fall into five general categories: raw data, derived data, documentation, reports, and administrative records (Table 1.1).

Table 2.1. Categories of data products and project deliverables.

Category	Examples
Raw data	GPS rover files, raw field forms and notebooks, photographs and sound/video recordings, telemetry or remote-sensed data files, biological voucher specimens
Compiled/derived data	Relational databases, tabular data files, GIS layers, maps, species checklists, analyzed data
Documentation	Data collection protocols, data processing/analysis protocols, record of protocol changes, data dictionary, FGDC/NBII metadata, data design documentation, quality assurance report, catalog of specimens/photographs
Reports	Annual progress report, final report (technical or general audience), periodic trend analysis report, publication
Administrative records	Contracts and agreements, study plan, research permit/application, other critical administrative correspondence

To meet our program goals and to ensure adequate context for the primary data products, the categories of project deliverables listed in Table 1.1 each require some level of management. This context in turn ensures the quality and availability of the primary data products. It is our intent to integrate the products listed in Table 1.1 such that project context within the program is maintained. This requires a holistic view of how natural resource data are generated, processed, finalized and made available. Such an understanding allows us to tailor our data management strategies to meet our objectives in the most efficient manner possible. Chapter II explores more thoroughly the types of data and information managed by the CAKN and the general process work flow by which these are generated.

I.3. Plan Scope

The data management plan stipulates the broad principles for handling ecological data from their most elementary state to final condition in reports or other dissemination mechanisms. While this plan places heavy emphasis on data generated by, or considered important to, the CAKN monitoring mission, the management principles presented may be applied to any data-gathering effort, short- or long-term, made by park resource staff. Core data management elements addressed in this plan include data collection, backup, quality control/assurance, validation, verification, documentation, archival, standard manipulation/analysis, organization, availability/access and distribution.

The procedures, strategy and guidelines put forth in this plan are intended for use in the management of any data the CAKN deems important to its mission.

Document Overview:

- Chapter I: Introduction, background information, plan overview
- Chapter II: Overview of the I&M data management process. CAKN-specific aspects, including vital sign monitoring components, are addressed in context of the overall I&M Program.
- Chapter III: Outline of the data management personnel, roles and responsibilities. Where possible, shared responsibilities are noted.
- Chapters IV through X: More specific data management infrastructure, standards and practices such as data acquisition, quality control, and documentation.
- Appendices: Information that directly supports this document as well as supplementary material supporting data management operations in general. The latter includes standard operating procedures for transferring program data to nationally maintained I&M databases such as NPSpecies, NatureBib and NR-GIS. It is anticipated that the appendices will experience the brunt of revisions, updates and additions in the data management plan both as the monitoring program develops and as general Program operations persist over time.

I.4. Plan Objectives

The goal of the NPS I&M Program is to provide scientifically and statistically sound data to support management decisions for the protection of park resources. The goal of data management is to ensure the quality, interpretability, security, longevity and availability of ecological data and related information resulting from resource inventory and monitoring efforts. The goal of the Data Management Plan is to outline the procedures and work practices that support effective data management. The DMP also serves as guidance for current and future staff of the CAKN, to ensure that sound data management practices are followed in any new data gathering efforts conducted or otherwise administered by the CAKN. Additionally, the DMP serves, where appropriate, to guide the enhancement of legacy data to match formats and standards put forth in this plan.

It is the intention of the CAKN I&M Program to establish sound data management practices for its monitoring mission such that they may be easily adopted by any resource management project. Further, where appropriate, the CAKN will promote effective data management and seek to educate personnel involved with the collection of resource data in the parks. By establishing robust data management, and maintaining the transparency of its inner workings, the CAKN hopes to lead by example and positively affect all data management within the parks.

It is the intention of the CAKN I&M Program to establish sound data management practices for its monitoring mission such that they may be easily adopted by any resource management project. Further, where appropriate, the CAKN will promote sound data management and seek to educate personnel involved with the collection of resource data in the parks.

Data management plan objectives include the following:

Overall Objectives:

- Outline the long-term goals of a comprehensive data management strategy for the CAKN I&M Program
- Associate data management goals with the long-term goals of the network and service-wide I&M program
- Outline the procedures and work practices that support effective data management
- Guide current and future staff of the CAKN to ensure that sound data management practices are followed
- Guide the enhancement of legacy data to match formats and standards put forth in this plan
- Encourage effective data management practices as an integral part of project management so all data are made available and usable for park management decisions now and into the future
- Optimize and promote interagency sharing and development of data, software applications and analyses

Specific Objectives:

- Establish roles and responsibilities of program staff for managing data
- Identify necessary elements for a functional data management program and describe any anticipated changes to those elements
- Establish an organization schema for Program data and information so that they are retrievable by staff, cooperators and the public
- Establish basic quality control standards
- Establish standards for data, data distribution and data archiving to ensure the long-term integrity of data, associated metadata and any supporting information

I.5. Plan Revisions

Appendix B outlines the standard revision schedule for this document. The plan will be formally reviewed and revised at least every 5 years. Informal review and revision will be on going to meet the changing needs of the program. The latest version of the DMP will be available on the CAKN website (*www.nature.nps.gov/im/units/cakn/DataMgt.htm*) and will include a revision log as an appendix.

II. I&M Data Management Process and Workflow

To ensure successful implementation, acceptance and long-term effectiveness, the infrastructure, procedures and objectives of data management must conform to the framework of typical resource project operations as well as overall I&M Program operations.

The project process flow presented in Chapter II is idealized. It is recognized that this flow and related data management strategies and infrastructure may not apply in all situations; however, planning for good data management requires an understanding of how work is or will be conducted in the Network. Where that understanding is not possible because Program operations are still in development, this plan takes the liberty of defining the process work flow.

An understanding of the tangible and common data-related products and procedures is also necessary for successful implementation of this plan. It is recognized that the base operating procedures and array of anticipated products for the CAKN are still in development at the first draft of this plan (submitted 12/15/2004). To the extent necessary to meet the goals of this plan, those procedures and products are surmised in the context of data management.

II.1. General Project Work

To assure effectiveness, the data management process must be pervasive in the execution of a project. At an operational level, the process used to carryout any data-gathering project should be clearly defined and governed by protocols that include relevant data management elements.

A generalized project work flow model provides a needed framework for data management. While the work flow presented in this plan may not apply to all situations, it does address both the long- and short-term data collection efforts anticipated by the CAKN monitoring program.

A project is divided here into the following stages:

1. Project Initiation
2. Planning and Approval
3. Design and Testing
4. Implementation
5. Product Delivery
6. Product Integration
7. Close Out and Evaluation

Project stages and related activities are depicted in Figure 2.1. Appendix D offers details of anticipated deliverables, deliverable repositories, and references to data management guidance documents for each stage.

Table 2.1 lists project tracking systems developed or under development in the NPS. General information about each of these can be found in Appendix E.
Additional tracking databases or systems are in place in individual parks or the regional office in Alaska.

As part of its monitoring program, the CAKN will use a project tracking database (MS Access) developed for the I&M Program in Alaska. To the extent feasible and reasonable, information in this project tracking database will be mapped to nationally maintained systems as appropriate. The primary function of the CAKN project tracking database will be support of monitoring program coordination and annual reporting. This project-tracking database will be modified as necessary as the monitoring program develops.

STRATEGY
A project tracking system geared toward data management for the I&M Program in Alaska is under development and expected to be ready for use in Fall 2004. The CAKN will adopt this tracking database and modify it if necessary as the monitoring program develops.

Table 2.1. Project tracking systems in the National Park Service.

Tracking System	Primary Purpose	Status
Project Management Information System (PMIS)	Project Funding	Active
Research Permit Reporting System (RPRS)	Permitting	Active
Planning, Environment, Public Comment (PEPC)	Compliance	Planned (Fall 2004)
Resource Activity Management System (RAMS)	Holistic Tracking	Planned (2004)

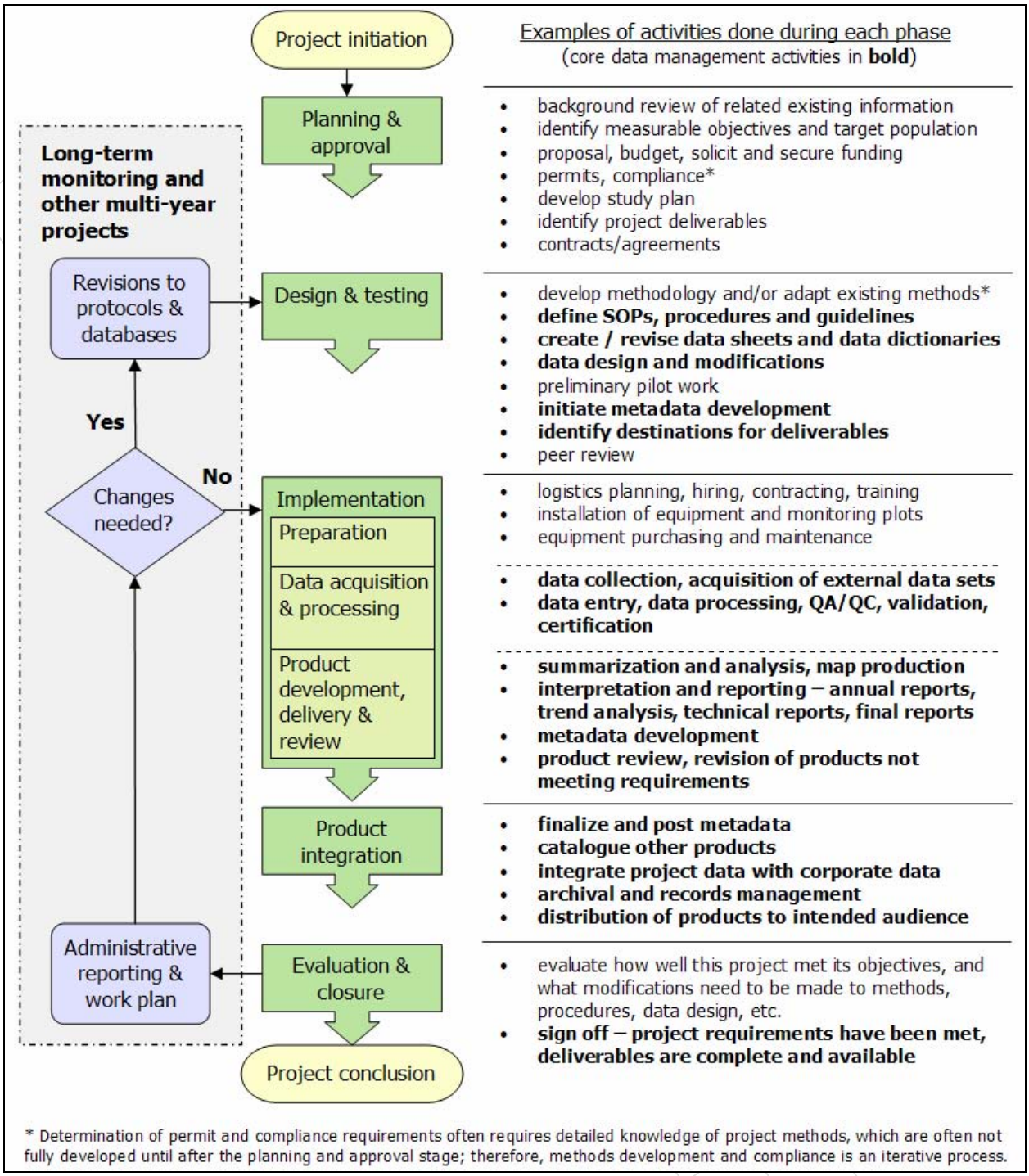


Figure 2.1. Generalized project stage flow for short- and long-term data collection efforts. Bold face activities indicate core data management elements.

Core Data Management Elements

Essential data and information management includes planning for data collection, actual collection, data entry, validation (quality control, etc.) and submittal to designated archives and audiences. Project activities involving these elements are highlighted in Figure 2.1.

As indicated in Figure 2.1, metadata should develop as the project progresses rather than after data are collected, entered, validated, verified and submitted. Similarly, quality control of data should begin with the data design. Field sheets and databases should be constructed to minimize, and where possible restrict, the variety of values that may be recorded for a given parameter. Pick lists, whether built into a data logger or included on a field sheet, should be employed wherever possible.

Certain key information is not only common to multiple data sets, but to the organization as a whole – lists of contacts, projects, parks, species are often complex and dynamic. It is a good strategy to centralize such information so that users have access to the most updated versions in a single, known place. Centralizing also avoids redundancy and versioning issues among multiple copies. Centralized information is maintained in database tables that can be linked or referred to from several distinct project databases. Network applications – for project management, administrative reporting, or budget management – can also link to the same tables so that all users in the network have instantaneous access to edits made by other users.

The CAKN will also utilize common tables for a variety of parameters collected by different vital sign projects. The national I&M Program's Natural Resource Table Standards (NRTS) is an effort to standardize commonly collected data such as percent of cloud cover, precipitation intensity, etc. and will provide a basis for common data values in the CAKN.

The National Park Service document "Recommended Database Strategies" also contains common data practices that the CAKN will adopt for vital sign monitoring. The document is available at *science.nature.nps.gov/im/datamgmt/links.htm*.

STRATEGY
The CAKN will utilize common pick lists for a variety of parameters collected by different vital sign projects. The national I&M Program's Natural Resource Table Standards (NRTS) is an effort to standardize commonly collected data such as percent of cloud cover, precipitation intensity, etc. and will provide a basis for common data values in the CAKN.

The National Park Service document "Recommended Database Strategies" also contains common data practices that the CAKN will adopt for vital sign monitoring. The document is available at *science.nature.nps.gov/im/datamgmt/links.htm*.

III. Data Management Roles and Responsibilities

Data management is about people and organizations as much as it is about information technology, database theories, and applications. Everyone within an organization uses or manages data and information. Thus, to serve the National Park Service and its constituents well, each of us within the Network must understand how data and information flow, and what our roles and responsibilities are in this process. A *role* is a function or position (*e.g. Data Manager*). A *responsibility* is a duty or obligation (*e.g. review data records*). This new and *crucial* emphasis on data management, analysis, and the reporting of results will require a large investment of personnel, time, and money. The CAKN Inventory and Monitoring Program expects to invest at least thirty percent of available resources in data management operations.

For the CAKN Inventory and Monitoring program to work effectively, everyone within the Network will have stewardship responsibilities in the production, analysis, management, and/or end use of the data. Table 3.1 summarizes the data stewardship roles of various Network personnel. Each of these broad categories has principal, or 'must-do', responsibilities as well as many potential ancillary tasks. As coordinator of these tasks, the fundamental role of the Network data manager is to understand and determine program and project requirements, to create and maintain data management infrastructure and standards, and to communicate and work with all responsible individuals.

The fundamental role of the Network data manager is to understand and determine program and project requirements, to create and maintain data management infrastructure and standards, and to communicate and work with all responsible individuals.

An effective data management plan specifies critical data management tasks, the personnel who execute them and any coordination between tasks and personnel needed to ensure efficient data management operations. The duties specified in the data management plan are constrained to those involving data management and do not represent the full range of duties for each role.

The remainder of this chapter discusses comprehensive data management roles and responsibilities that generally apply to all Network activities. Each Vital Sign monitoring protocol and any project study plan contains specific instructions for assignments and tasks that nest within this overall framework. Individuals who carry out monitoring protocols and inventory study plans are responsible for reading and understanding these instructional guidelines.

The remainder of this plan contains references to specific duties and obligations pertaining to selected data management topics. All of these, however, fall under the scope of roles and responsibilities covered in this chapter.

Table 3.1. Categories of data stewardship involving all Network personnel.

Stewardship Category	Related Activities	Principal jobs or positions
Note: Each position is listed in only one category according to overriding responsibilities. However, most positions contribute in each category.		
Production	Creating data or information from any original or derived source. This includes recording locations, images, measurements, and observations in the field, digitizing source maps, keying in data from a hardcopy source, converting existing data sources, image processing, and preparing and delivering informative products, such as summary tables, maps, charts, and reports.	Project Crew Member Project Crew Leader Data/GIS Specialist or Technician
Analysis	Using data to predict, qualify, and quantify ecosystem elements, structure, and function as part of the effort to understand these components, address monitoring objectives, and inform park and ecosystem management.	Network Ecologist Resource Specialist
Management	Preparing and executing policies, procedures, and activities that keep data and information resources organized, available, useful, compliant, and safe.	Network Coordinator Network Data Manager Project Leader GIS Manager Information Technology Specialist Database Manager National Level I&M Data Manager
End Use	Obtaining and applying available information to develop knowledge that contributes to understanding and managing park resources.	Network Coordinator Park managers and superintendents Others

III.1 Data Management Roles and Responsibilities

An increasing demand for more detailed, higher quality data and information about natural resources and ecosystem functions requires a group of people working together to steward data and information assets. Knowledgeable individuals from many areas must work in concert to ensure that data are collected using appropriate methods, and that resulting data sets, reports, maps, models, and other derived products are well managed. Data sets and the presentations of these data must be credible, representative, and available for current and future needs.

Table 3.2 summarizes the roles and responsibilities of various personnel. These roles are listed 'from the ground up' to help demonstrate the hierarchy of responsibilities. For example, a project leader is ultimately responsible for the activities listed in the field level

roles of crew leader and crew member. Also, the network coordinator ensures that the network data manager and ecologist achieve the required performance level. Appendix F reiterates and expands the roles and responsibilities summarized in Table 3.2.

Table 3.2. Summary of Roles and Responsibilities.

Role	Primary responsibilities related to data management
Project Crew Member	Record and verify measurements and observations based on project objectives and protocols. Document methods, procedures and anomalies.
Project Crew Leader	Supervise crew members to ensure their data collection and management obligations are met, including data verification and documentation.
Data/GIS Specialist or Technician	Perform assigned level of technical data management and/or GIS activities, including data entry, data conversion, and documentation. Work on overall data quality and stewardship with project leaders, resource specialists, and the Network data manager.
Information Technology/Systems Specialist (park IT support staff)	Maintain local area network, establish and maintain system security, and keep software and hardware systems up to date. Additionally, IT staff are responsible for maintaining connections between the LAN and the internet. They work with the Network data manager and GIS liaisons to establish a directory structure (including drive partitions) that provides local access and security for natural resource data. They manage the infrastructure for digital data backups for the local area network.
Network Coordinator	Interface with project leaders to ensure that timelines for data entry, validation, verification, summarization/analysis and reporting are met. Additionally, the Network Coordinator must review and approve proposed changes to project protocols prior to implementation.
Project Leader	Oversee and direct operations, including data management requirements, for one or more Network projects. Maintain communication with project staff, Network Data Manager, and resource specialist regarding data management. **Note:** The Project Leader is often a resource specialist, in which case the associated responsibilities for data authority apply (see next role). A Project Leader without the required background to act as an authority for the data will consult with and involve the appropriate resource specialists.
Resource Specialist	Understand the objectives of the project, the resulting data, and their scientific and management relevance. Make decisions about data with regard to validity, utility, sensitivity, and availability. Describe, publish, release, and discuss the data and associated information products. **Note:** The Resource Specialist serving as a Project Leader is also responsible for the duties listed with that role.
GIS Manager	Support park management objectives. Coordinate and integrate local GIS and resource information management with Network, Regional, and National standards and guidelines.

Role	Primary responsibilities related to data management
Network Data Manager	Provide overall Network planning, training, and operational support for the awareness, coordination, and integration of data and information management activities, including people, information needs, data, software, and hardware. Serve as Point of Contact for National Park Service database applications (NPSpecies) Coordinate internal and external data management activities.
Database Manager	Apply particular knowledge and abilities related to database software and associated application(s)
Network Ecologist	Ensure useful data are collected and managed by integrating natural resource science in network activities and products, including objective setting, sample design, data analysis, synthesis, and reporting
I&M Data Manager (National Level)	Provide service-wide database design, support, and services, including receiving and processing to convert, store, and archive data in service-wide databases
Other End Users	These 'information consumers' include park managers and superintendents, researchers, staff from other agencies, and the public. End users are responsible for the appropriate use and application of data and derived products and for providing feedback for improvements.

III.2 Project Stewardship

Since the data management aspects of every inventory or monitoring project normally require the expertise and involvement of several people over a period of months or years, it makes sense that one person is charged with keeping track of the objectives, requirements, and progress for each project. This project leader (or *steward*) is usually a resource management specialist with training and experience in the field of science related to the inventory or monitoring project. The project leader must act as a steward for the data and work with Network and project personnel to ensure the proper handling of data for the project.

Data Stewardship–Sharing Responsibilities

Successful data stewardship requires that people involved in CAKN activities learn and understand Network expectations for continuous data management. This is equally important for Network staff, park employees, and contractors or cooperators. All project participants receive training, briefings, materials, and additional regular communication about data stewardship from supervisors, project leaders, and data managers. The purpose is to promote the appropriate level of understanding about how their efforts relate to park and network management objectives, National Park Service and Department of Interior policies, and other federal government requirements. Other relevant context and linkages can also be discussed to help establish a sense of ownership and accountability with project staff.

Inventory and Monitoring project leaders must understand resource information management issues and requirements, and they must be aware of the challenges and limitations of field data collection, including details like the use of GPS. We can achieve these goals through thorough documentation, detailed and regular briefings and by preparing well-trained field crews to collect data at reasonable intervals.

The Importance of Documentation

If one shared responsibility stands above the pack in importance and value, it is the careful documentation of data sets, the data source(s), and the methodology by which the data were collected or acquired. This documentation establishes the basis for the appropriate use of the data in resulting analysis and products, both in the short-term and long-term. Network monitoring protocols contain key elements of data documentation. Network data records collected according to these protocols will include the name, date, and version of the associated protocol. (Chapter VII presents important guidance and reference for documentation and metadata)

The Hub of Data Stewardship

Network coordinators, project leaders and data managers (and potentially GIS specialists and other specialists) comprise the central data management team for a given inventory and monitoring projects. Each is responsible for certain aspects of project data and all share responsibility for some overlapping tasks. Because of the collaborative nature of project data management, good communication among these personnel is essential to meeting program goals.

Network Coordinator

The Network Coordinator interfaces with project leaders to ensure that timelines for data entry, validation, verification, summarization/analysis and reporting are met.

Project Leader

Project leaders oversee and supervise all phases of an inventory and monitoring project from initiation to product delivery. They act as the point of contact for that project and are responsible for complying with program policy, project protocol methods and the network data management plan. The project leader is responsible for final submission of all products and deliverables. Project leaders work with the Network data managers to ensure that proper metadata are generated for project data and are submitted along with final products to national program databases (including NPSpecies, NR-GIS, NatureBib and ANCS+). Their active involvement in data management determines the quality and usefulness of the project data and overall success and longevity of the I&M Program.

In terms of data management, the project leader assumes “ownership” of the data meaning that as leader of the data-gathering effort (project). In terms of the CAKN monitoring program, each vital sign is considered a project and will be assigned leadership by the Network technical committee.

To ensure the quality of each project, including data requirements, project leaders are responsible for designating an alternate leader who is capable of maintaining project operations in his or her absence.

STRATEGY
Project leaders are responsible for designating an alternate leader who is capable of maintaining project operations in his or her absence.

Specifically, a project leader is responsible for:

- *project documentation* that describes the ‘who,’ ‘what,’ ‘where,’ ‘when,’ ‘why,’ and ‘how’ of a project
- documentation and implementation of *standard procedures* for field data collection and data handling

- quality assurance and *quality control measures*, which include the supervision and certification of all field operations, staff training, equipment calibration, species identification, data collection, data entry, verification, and validation
- maintenance of concise explanatory documentation of all deviations from standard procedures
- detailed documentation for each field data collection period
- maintenance of hard copies of data forms and archiving of original data forms
- scheduling of regular project milestones such as data collection periods, data processing target dates, and reporting deadlines
- regular summary reports, periodic trend analysis of data, resulting reports, and their public availability
- identify sensitive information that requires special consideration prior to distribution
- acting as the main point of contact concerning data content/quality

The project leader may also work closely with the data manager and/or a data specialist (such as a biometrician) to:

- develop quality assurance and quality control procedures specific to project operations
- identify training needs for staff related to data management philosophy, database software use, and quality control procedures
- coordinate changes to the field data forms and the user interface for the project database (coordinate data entry procedures – data design maintenance)
- document and maintain master data (including metadata generation and maintenance)
- manage the archival process to ensure regular archival of project documentation, original field data, databases, reports and summaries, and other products from the project
- define the process of how project data will be transformed from raw data into meaningful information
- create data summary procedures to automate and standardize this transformation process
- identify and prioritize legacy data for conversion and convert priority data sets to a modern format
- increase the interpretability and accessibility of existing natural resource information
- catalogue project data and reports in nationally maintained I&M Program databases such as NPSpecies, NatureBib and NR-GIS (see Appendix C).

Network Data Manager

The data manager oversees the development, implementation, and maintenance of data infrastructure and standards for the Network. The data manager facilitates coordination between projects and protocols to allow for interchange of data and information where possible. The data manager also facilitates the long-term storage and maintenance of Program output. The data manager works with project leaders and other Program staff to design databases, applications, and products, and to facilitate dissemination of both project data and products.

The network data manager exercises stewardship over the data to ensure data are archived, documented, made compatible with other program data and made

available/discoverable. Further, the data manager ensures that the mechanisms in place for making data discoverable and available are maintained and, along with the project leader, that the information conveyed via these mechanisms is up-to-date and accurate.

The I&M Program Network Data Manager position description is included as Appendix G.

General data management duties for the data manager are:

- overall coordination of data management activities in the Network
- develop and maintain data and information housing and dissemination system for all program data and information
- work to improve the acquisition, accessibility and transparency of digital data
- ensure that the data and information system is populated and kept up-to-date with all relevant Program output
- develop and maintain logs recording both the changes/enhancements to the data and information-handling system/process and use/dissemination of system content (actual data and information)
- act as point of contact for access to Program output
- ensure data security (archiving operations, etc.)

Data managers will also work closely with the project leader to:

- develop and maintain the infrastructure for metadata creation, project documentation, and project data management
- create and maintain project databases in accordance with best practices and current program standards
- provide training in the theory and practice of data management tailored to the needs of project personnel
- establish and implement procedures to protect sensitive data according to project needs
- collaborate with GIS specialists to integrate tabular data with geospatial data in a GIS system in a manner that meets project objectives
- define the scope of the project data and create a data structure that meets project needs
- become familiar with how the data are collected, handled, and used
- review quality control and quality assurance aspects of project protocols and standard procedure documentation
- identify elements that can be built into the database structure to facilitate quality control, such as required fields, range limits, pick-lists and conditional validation rules
- create a user interface that streamlines the process of data entry, review, validation, and summarization that is consistent with the capabilities of the project staff
- develop automated database procedures to improve the efficiency of the data summarization and reporting process
- make sure that project documentation is complete, complies with metadata requirements, and enhances the interpretability and longevity of the project data
- ensure regular archival of project materials
- inform project staff of changes and advances in data management practices

Figure 3.1 illustrates data management responsibilities for project leaders and Network data managers.

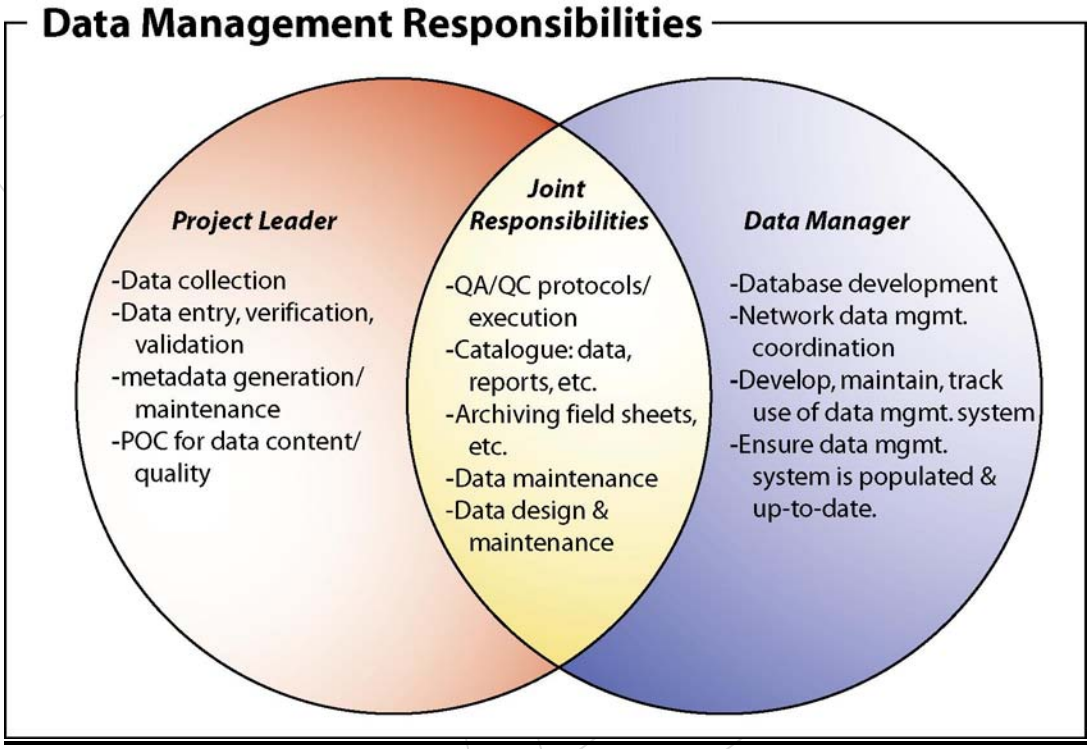


Figure 3.1. Basic data management roles and responsibilities for project leaders and Network data managers.

AKRO GIS Liaisons (GIS Specialists)

The GIS specialists manage spatial data themes associated with network inventory and monitoring projects, as well as other spatial data related to the full range of park resources. They incorporate spatial data into the GIS. They also maintain standards for geographic data and are responsible for sharing and disseminating GIS data throughout the network. As noted in Chapter II, the CAKN maintains an agreement with the AKRO GIS Team for Network GIS needs.

The GIS specialists work in collaboration with project leaders to:

- determine the GIS data and analysis needs for the project
- develop procedures for field collection of spatial data including the use of GPS and other spatial data collection techniques
- display, analyze, and create maps from spatial data to meet project objectives
- properly document data in compliance with spatial metadata standards

GIS specialists will also work directly with data managers to:

- design databases and other applications for the network
- create relationships between GIS and non-spatial data
- create database and GIS applications to facilitate the integration and analysis of both spatial and non-spatial data

- establish and implement procedures to protect sensitive spatial data according to project needs
- develop and maintain an infrastructure for metadata creation and maintenance
- ensure that project metadata are created and comply with national and agency standards

Project needs are articulated to the appropriate GIS liaison either by the project leader or the data manager at the earliest possible stage in the development of a given protocol.

STRATEGY
CAKN monitoring program GIS needs will be articulated to the AKRO GIS team at the earliest possible stage in the development of a given protocol. Currently, the CAKN has one liaison assigned for each park in the network. In general, the CAKN will communicate GIS needs to all three liaisons in unison, depending on the intended implementation of a given protocol.

Currently (12/15/2004), the CAKN has one liaison assigned for each park. In general, the CAKN will communicate GIS needs to all three liaisons in unison, depending on the intended implementation of a given protocol. The AKRO GIS review committee will approve, assign and schedule tasks as needed.

III.3 Data Management Coordination

The Natural Resource Challenge states that collaboration among the National Park Service, other public agencies, universities, and non-governmental organizations is necessary to effectively acquire, apply, and promulgate the scientific knowledge gained in National Parks. The I&M Program encourages coordination among participants at all levels to help ensure that data collected by NPS staff, cooperators, researchers, and others are entered, quality-checked, documented, analyzed, reported, archived, cataloged, and made available for management decision-making, research, and education. The CAKN Monitoring Plan discusses in more detail partnerships with other organizations or agencies.

Selected Data and Document Communities in Alaska:

- Alaska Resources Library and Information System (ARLIS)
- Alaska Geospatial Data Committee (AGDC)
- Alaska Depart of Fish and Game
- Alaska Depart of Natural Resources
- Alaska Biological Information Consortium (ABIC)
- University of Alaska
- U. S. Bureau of Land Management
- U. S. Fish and Wildlife
- USGS

Alaska Data Management Work Group

The NPS Alaska data management work group consists of the four network data managers and selected regional support staff. The regional support staff represent: tabular data management, GIS data management, field work data management, and programming. This work group works toward increasing the awareness of good data management among park staff, resource scientists, and others who produce, maintain or use ecological data and related information. Additionally, the data management work

group ensures that Program output across the region is compatible and handled in a common fashion. The workgroup ensures that data and information are managed across the Network and Region such that all components may be easily compared by location, time and subject. Supporting documentation (metadata, analysis reports, etc.) will also be produced in a consistent manner across the Network and Region.

The data management workgroup will work to ensure that data and information across the Network and Region may be easily compared by location, time and subject.

The Alaska I&M Data management workgroup works to:

- Promote the awareness of good data management among park staff, resource scientists, and others who produce, maintain or use ecological data and related information
- Ensure that data and information are managed across the networks and Region such that all components may be easily compared by location, time and subject
- Ensure that supporting documentation including metadata, reports, etc. are produced in a consistent manner across the networks and Region
- Produce tools for data acquisition, validation, summary, analysis, access and distribution using state-of-the-art information technology including custom desktop and web applications

The Network data manager works with I&M Program data management staff and regional resource information management personnel to maintain a high-level of involvement in service-wide and regional databases and data management policy. The Network data manager also works locally with Network personnel, park staff, and cooperators to promote and develop workable standards and procedures that result in the integration and availability of datasets.

Key contacts for the Network data manager include:

- Park GIS and data managers
- Project leaders for each monitoring or inventory project
- National-level I&M information managers
- Other Network data managers in the I&M Program

Consistent and productive communication among these personnel leads to common understanding and better synchronization of Network and park data management activities. Involvement and input from Park scientists and resource information management staff is essential. Everyone within the Network has a role in the successful development of the materials that define our work including planning documents, monitoring protocols, populated datasets and reports.

<u>Network of Networks</u>

Data managers throughout the I&M Program regularly coordinate with each other and national program staff via annual meetings, conference calls, workgroups, a listserv, web sites, and informal communication. Data managers from these networks share the workload and work with each other to develop their respective Network Data Management Plans. This model of cooperation and communication is effective, and it can be applied to resource information management issues and used by administrators.

The CAKN maintains an active role in promoting practical consistency among protocols and data sets involving other networks and organizations. GIS and data managers from all Network parks are encouraged to participate in program development and activities, to use I&M Program resources, and to communicate with Network and Program staff to share information about progress and direction and to address concerns.

The CAKN maintains an active role in promoting practical consistency among protocols and data sets involving other networks and organizations.

IV. Data Management Infrastructure

The CAKN computer resource infrastructure is composed of computers and servers that are functionally or directly linked through computer networking services. This infrastructure represents the foundation upon which our Network information system is built. System architecture signifies the applications, database systems, repositories, and software tools that make up the framework of our data management enterprise.

The CAKN monitoring program relies heavily on park, regional and national information technology (IT) personnel and resources to maintain the computer resource infrastructure. This includes but is not limited to hardware replacement, software installation and support, security updates, virus-protection, telecommunications networking, and backups of servers. Therefore communication with park and regional IT specialists is essential to ensure adequate resources and service continuity for our systems architecture. Rather than focusing on a detailed description of a snapshot of our current computer resources, this chapter will instead describe our infrastructure in more general terms and focus more specifically on the systems architecture that is central to data management.

The CAKN data management "system" is currently (12/15/2004) in development. The system design presented here represents current plans which may be significantly altered by the point of full implementation.

This chapter also specifies standards for CAKN data management operations. In general, the CAKN will conform to National Park Service standards and policy in all aspects of Program data management operations. Further, the CAKN will conform to national I&M Program standards and mandates in the interest of program integration and information sharing.

IV.1. NPS Infrastructure - Components

An important element of a data management program is a reliable, secure network of computers and servers. Our digital infrastructure has three main components: park-based local area networks (LAN), network data servers, and servers maintained at the national level. This infrastructure is maintained by park, regional, and national IT specialists, who administer all aspects of system security and backups. Figure 4.1 illustrates these components.

The information system architecture necessary to fulfill the role of Program data management will include both existing and planned components. National-level I&M data management infrastructure and strategy is presented in Appendix C. Existing Regional-, Network- and Park-level infrastructure will be augmented with additional components required in this plan to meet Program data management objectives.

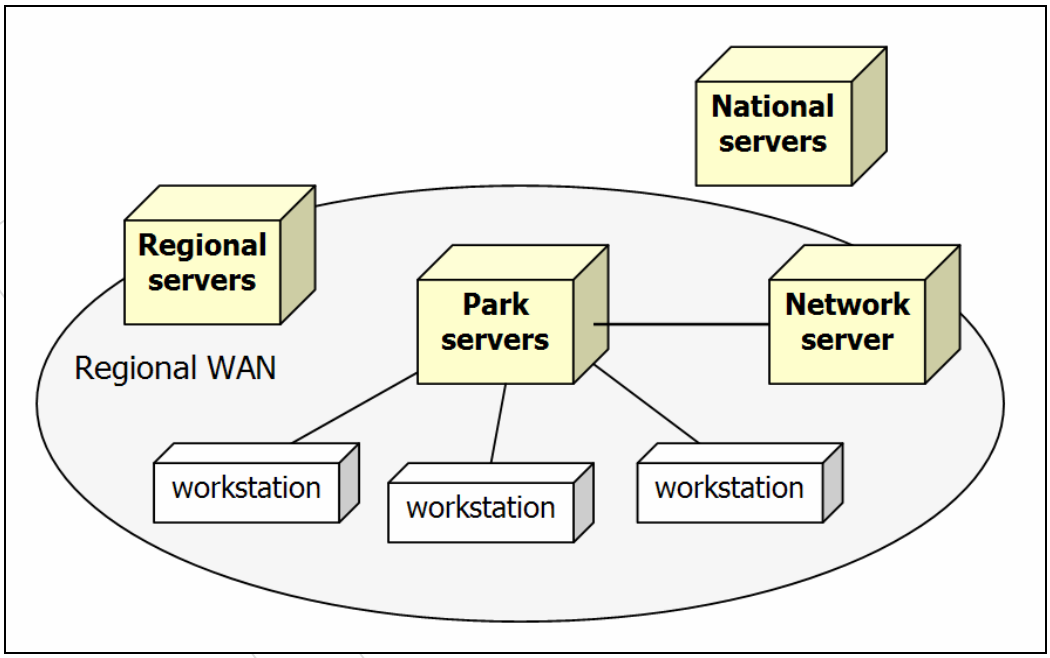


Figure 4.1. Principle information technology components.

Regional-Level Infrastructure

The Alaska Regional Office (AKRO) offers a central repository for many types of information and acts as the backbone for base cartographic data in the region. Further, the regional office maintains a wide area network file server for general file exchange and storage. The AKRO is also actively working to improve regional computer networking and setup regional-level client server database architecture for use by various NPS programs including the I&M Program.

The AKRO GIS Team provides support of various spatial data operations via a request process and a liaison assigned to each park. A general understanding of support exists between the CAKN and the GIS Team and agreements between the two be established for vital sign measures requiring significant and/or annual GIS support.

STRATEGY
The CAKN will establish an agreement with the AKRO GIS Team for vital sign measures requiring significant and/or annual GIS support.

The AKRO GIS Team also maintains a server that will act as offsite backup for all CAKN data. Implementation of this plan will include arrangement for transfer of CAKN data to the AKRO GIS server.

STRATEGY
The AKRO GIS Team maintains a server that will act as offsite backup for all CAKN data.

To summarize, the AKRO provides the following that the CAKN will utilize to meet its goals:

- The Wide Area Network file server for general file exchange and storage
- GIS and related tabular data accessible via custom applications distributed to the parks as well as the Alaska GIS Data Clearing House (*www.nps.gov/akso/gis*).
- File server to provide offsite storage for all CAKN data

Network-Level Infrastructure

The CAKN infrastructure, that is the computers and personnel used to carry out the CAKN mission, is composed of elements from the national I&M Program, the NPS Alaska region and CAKN-member parks (Figure 4.1). For general operations, the CAKN relies primarily on the system infrastructure of the Yukon-Charley Rivers and Gates of the Arctic (YUGA) National Park office. Regional infrastructure elements are noted above. The Network utilizes the national infrastructure outlined in Appendix C for its public website presence and general file-sharing via the NPS FTP site.

The CAKN infrastructure, that is the computers and personnel used to carry out the CAKN mission, is composed of infrastructure elements from the national I&M Program, the NPS Alaska region and CAKN-member parks.

The common directory structure the CAKN will use for file storage is shown in Figure 4.2. The "corporate" directory structure will be maintained on the primary CAKN repository in the YUGA office. Duplicate structures will be established on local drives at each park (see Park-Level Infrastructure below).

STRATEGY
The "corporate" directory structure will be maintained on the primary CAKN repository in the YUGA office. Duplicate structures will be established on local drives at each park.

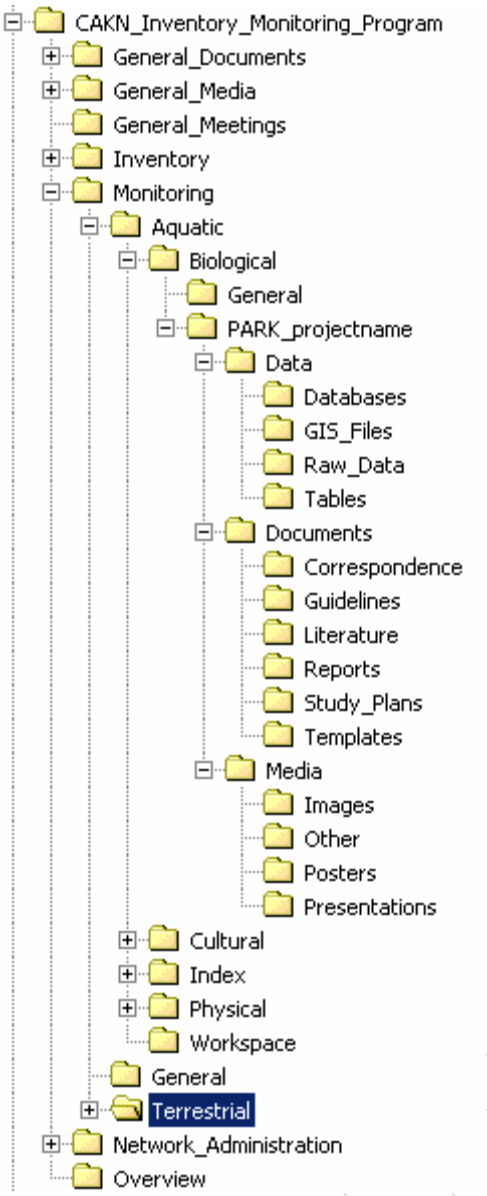


Figure 4.2. File directory structure. Folders in this example are expanded for a generic aquatic monitoring project.

Chapters IX and X of this plan present more details on the distribution, maintenance and storage of CAKN data. These chapters may be summarized as:

- The CAKN will manage a primary repository located in the YUGA office for data and information generated by the Network. These data will be accessible via custom applications as well as the CAKN website and open to authorized NPS personnel.
- The primary CAKN repository will be backed up to an offsite server in the AKRO.
- Required portions of the CAKN data will be stored in national I&M Program databases such as NPSpecies, NatureBib and NR-GIS.

- Certain CAKN datasets will be maintained by outside organizations however metadata for these will be maintained in the primary CAKN repository. An example of this category is the climate data which will be handled by the Western Regional Climate Center under formal agreement with the CAKN.

Park-Level Infrastructure

Each park maintains a local area network (LAN) with at least one shared server for employee use. Currently (12/15/2004), CAKN documents and files are stored on park-network drives at the discretion of the various project leaders assigned to different pilot projects. However, a directory structure mirroring the regional and national recommendations is in use on the YUGA LAN.

As noted above, directory structures mirroring that of the primary CAKN repository will be established at each member park. In accordance with current (12/15/2004) park office LANs, these structures will reside on the following local drives:

- DENA – T:\CAKN_Inventory_Monitoring_Program
- WRST – T:\CAKN_Inventory_Monitoring_Program
- YUCH – K:\CAKN_Inventory_Monitoring_Program

Park-based local areas networks will serve as connections to local file servers housing working databases for initial archival of raw data, data entry and data cleansing. Regional network connections will serve to transfer these working databases for upload to server-based data stores in Fairbanks and subsequently Anchorage for offsite storage. These server-based databases will serve CAKN I&M data and information to the NPS Alaska region. WASO servers will make available the portion of the program output deemed fit for public consumption via NPSpecies, NatureBib and NR-GIS (see Appendix C).

In general, CAKN operations are designed to function within and augment Park operations (Figure 4.3).

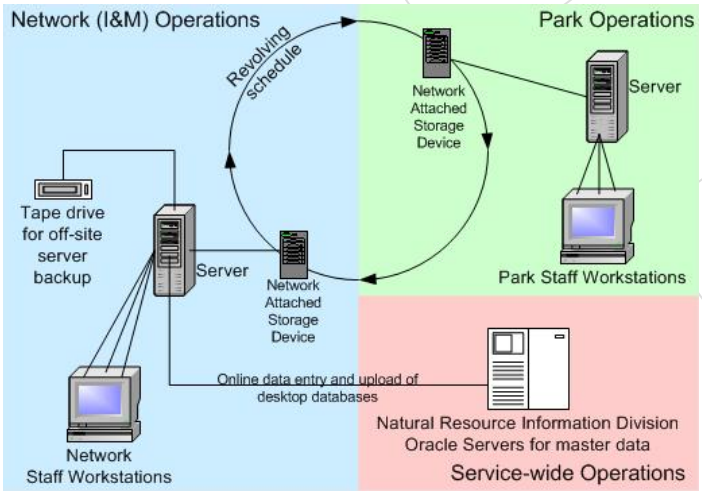


Figure 4.3. Information technology connectivity between Network, Park and National operations.

STRATEGY
CAKN Park-Based "Team" Drives for I&M:
DENA – T:\CAKN_Inventory_Monitoring_Program\
WRST – T:\CAKN_Inventory_Monitoring_Program\
YUCH – K:\CAKN_Inventory_Monitoring_Program\

Park-based local areas networks will serve as connections to local file servers housing working databases for initial archival of raw data, data entry and data cleansing. Regional network connections will serve to transfer these working databases for upload to server-based data stores in Fairbanks and subsequently Anchorage for offsite storage. These server-based databases will serve CAKN I&M data and information to the NPS Alaska region. WASO servers will make available the portion of the program output deemed fit for public consumption via NPSpecies, NatureBib and NR-GIS.

IV.2. Data Management Hardware Standards

CAKN information management hardware will conform to NPS standards and policy. In practice, the majority of hardware configurations will be subject to existing park-based IT infrastructure. Where possible and appropriate, the CAKN will take advantage of new developments in computer technology to enhance data collection, quality control, storage and accessibility.

IV.3. Basic Software Standards

The software used by the CAKN will conform to NPS standards and policy. The subsections below provide further detail regarding standard practices within the CAKN.

Word Processing

Microsoft Word and Adobe Acrobat. All reports and other textual documents will be finalized in electronic format in the latest available version of MS Word unless otherwise specified in advance (in the project study plan). Distribution copies will be converted to the latest version of Adobe Acrobat.

Databases

Desktop: Microsoft Access. Desktop versions of CAKN databases will be in the latest MS Access format unless otherwise specified in advance (in the project study plan). Any database work done under contract will also use MS Access unless otherwise specified in advance.
Client/Server: Microsoft SQL Server. The CAKN is in the process of developing a server-based relational database as a core repository for Program output (see Chapters IV and IX for additional information).

GIS Products

ArcGIS. All GIS products must be compatible with ESRI's ArcGIS software. GIS products should meet NPS specifications

(*www.nps.gov/gis/data_standards/DataStandards.html#Stewardship*) and where appropriate will be archived as in uncompressed TIF and PDF files. All GIS products will be accompanied by FGDC-compliant metadata. Chapter IX specifies additional details for the dissemination of program data.

IV.4. Data Management System

Section IV.1 presents an overall project process flow as a framework for data and information management. Core date management practices are highlighted and put in context of the project process flow in Figure 2.1. The major elements in the process for executing data management within the framework may be summarized as:

- Acquire data (field work, satellite download, etc.)
- Raw data archived (physical and possibly digital material)
- Entry/import
- Verification/validation (quality assurance and control)
- Documentation (metadata, etc.)
- Validated data & documentation archived (digital)
- Cataloguing ("integration": making content broadly discoverable: NR-GIS, NatureBib, etc.)
- Analysis & reporting
- Analyzed/reported data archived (products archived)
- Analyzed/reported data catalogued

These elements embody the core of any data management system in the I&M Program. For that system to be useful in meeting the goals of the I&M Program, the ability to access, view and comprehend as well as extract and manipulate, the content (data and information) must be systemic to that system. This ability is typically manifested in a software application, be it web-based or not, designed as a window to system content. While this data management plan does not address the details of such an application, Section IV.5 outlines the core functionality that the CAKN seeks in such a tool. Further, we specify within the context of overall system design, how, at least initially (12/15/2004), the data will be accessed, viewed, and extracted.

The elements listed above roughly comprise a data life cycle within the CAKN. Figure 2.1 presents the relative timing of these elements within a project. For the life cycle to have actionable meaning, we must also specify how and where these core data management elements will be done. For the most part, these are documented in the standard operating procedures for each vital sign. There are, however, some commonalities to these procedures dictated in this plan.

Figure 4.4 presents a generalized schematic of the data life cycle for CAKN data. While variations to the cycle occur based on project specifics (e.g. in the case of climate or game harvest data which will be managed by outside entities), the majority of data handled by the CAKN will follow this general course.

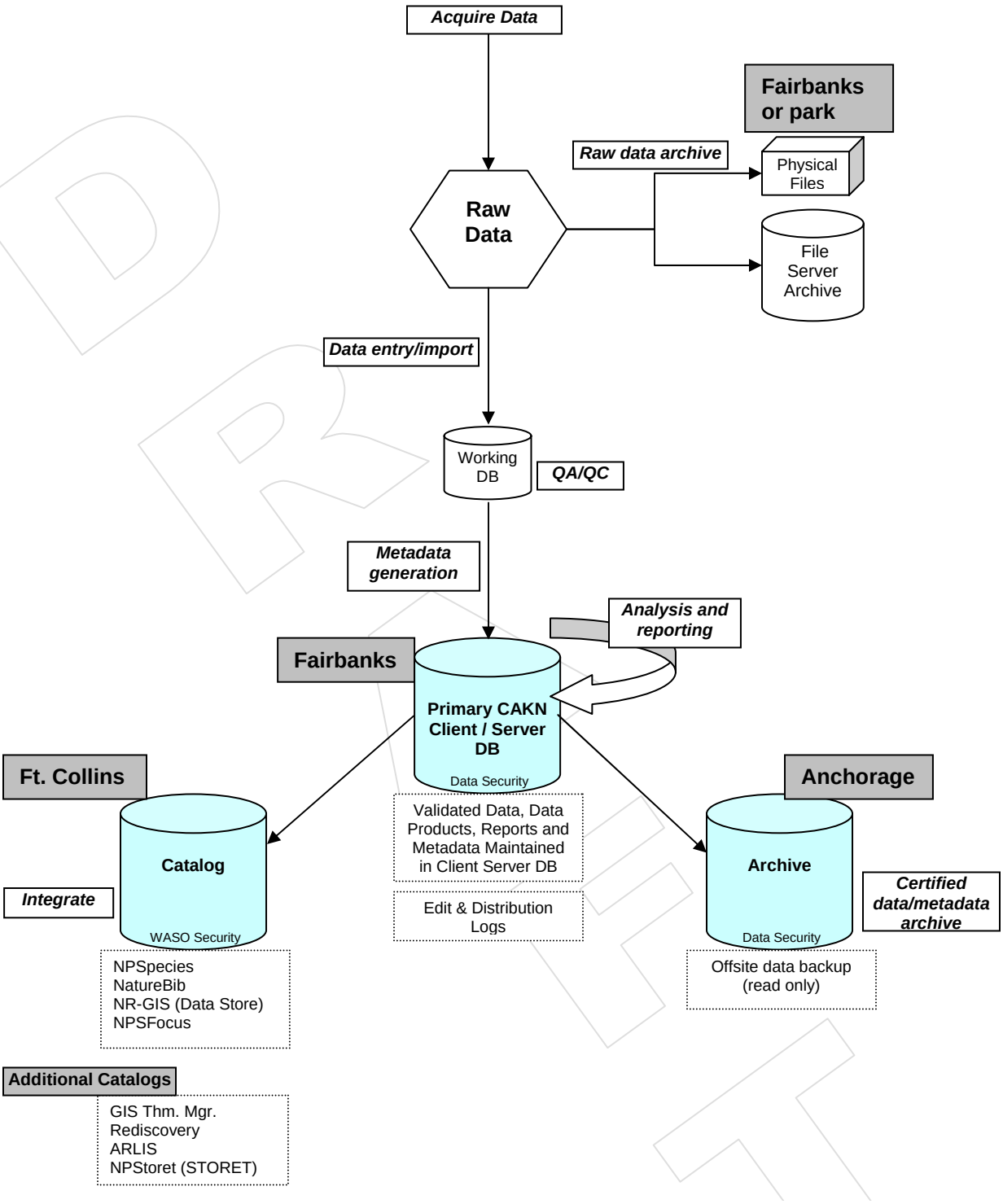


Figure 4.4. Generalized path of data within the CAKN. The primary data server is accessed via desktop and web-based applications. Analyzed or derived data products are stored on the primary server, archived and catalogued (integrated) along with certified data.

IV.5. Data Access Applications – Getting at the data and information

It is impossible to predict wholly the demands which will be made on a data interface for the complex body of data and information the CAKN intends to manage. We assume, however, the following basic needs:

- Robust data browsing:
    - A quick way to see summary results from any one vital sign measure
    - A quick way to compare results between vital signs
    - A quick way to view monitoring results by time and location
- Robust data access (download):
    - Easy data subsetting (data searches that allow data to be selected by data, location, subject, etc.)
    - Useful format choices
    - Dissemination tracking
    - Accompanying metadata
- Robust data upload for Project Leaders and authorized personnel
- Robust data discovery so that our data and information may be easily found by outside stake holders

V. Data Acquisition and Processing

The National Park Service's Natural Resource Inventory and Monitoring Program, in support of the NPS Natural Resources Challenge, is responsible for acquiring the information required by park managers to effectively manage and maintain park natural resources. To successfully accomplish this task, information from multiple sources is collected by the CAKN I&M Program (Table 5.1) and processed so that it meets national, regional and network standards. Chapter VI addresses quality control and assurance regarding the collection of data. This chapter presents an overview of the data important to the CAKN mission with special attention to monitoring data. The general strategy for when monitoring data will be acquired and where it will go is addressed. Also addressed is the acquisition of data from non-NPS programs as well as from data mining.

Data acquisition begins with data design (data collection SOPs and database structure) and is significantly impacted by choices made early in project planning. Data designs for all CAKN monitoring data and CAKN-initiated projects will conform to the I&M Natural Resource Template Standards (formally the "Buffet of Fields and Tables") and confine, where possible, data values to specific ranges. Data acquired by the CAKN will be developed, if necessary, as digital datasets conforming to NPS and I&M database standards.

STRATEGY
Data designs for all CAKN monitoring data and CAKN-initiated projects will conform to the I&M Natural Resource Template Standards and confine, where possible, data values to specific ranges.

Typically, data are acquired via field work but may also be obtained from sources outside the NPS. Table 5.1 summarizes the data to be collected by the CAKN monitoring program. Upon the completion of any field work, a summary of what was collected will be entered in the project tracking database (see Section II.1). Acquisition of data sourced outside the NPS will be addressed in the appropriate Vital Sign protocol.

STRATEGY
Upon the completion of field work, a summary of what was collected will be entered in the project tracking database.

V.1. CAKN Data

Data and information for the CAKN may be organized into five categories: Monitoring, Inventory, Prototype (DENA LTEM), Short-term (e.g. pilot projects) and Legacy (primarily "mined" data). Together these broadly constitute the data and information universe for the CAKN (Figure 5.1). Each of these falls within either the "Raw" or "Compiled/derived" categories shown in Table 1.1.

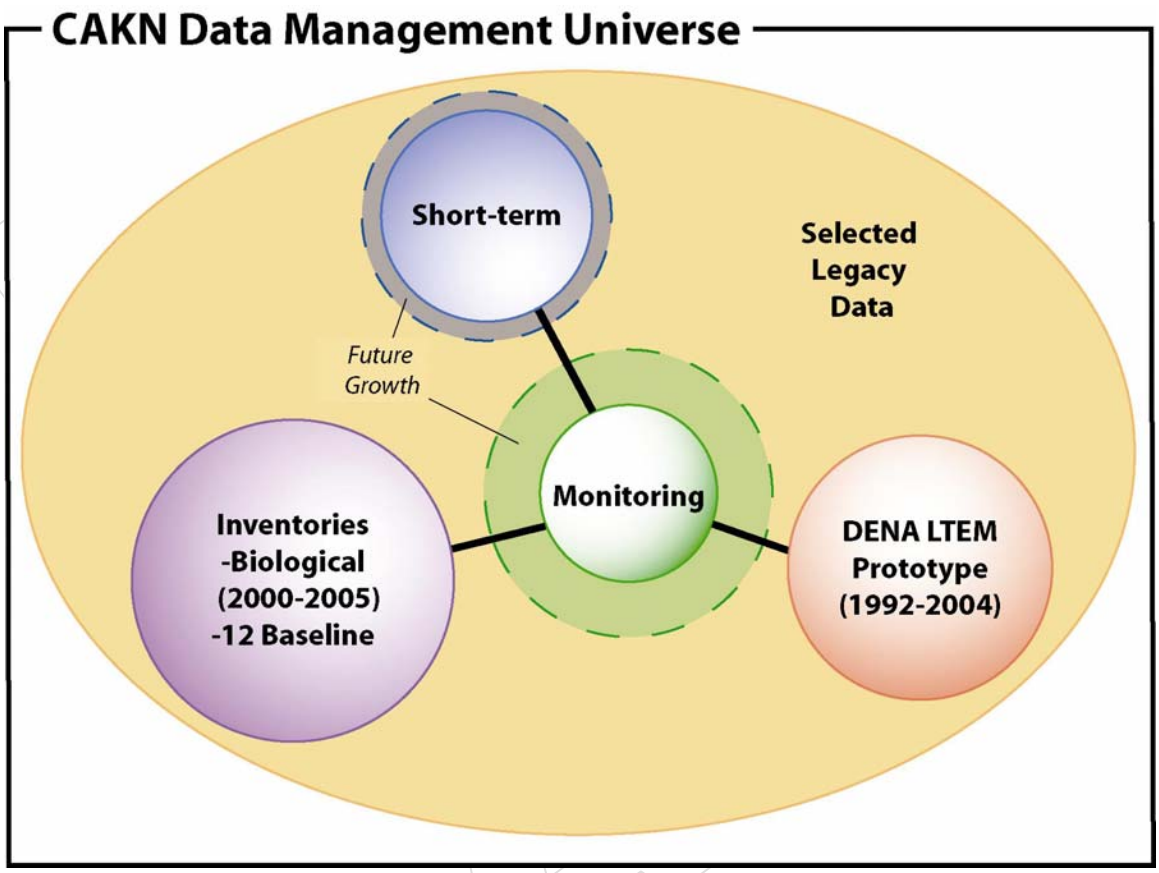


Figure 5.1. Broad categories of data for the CAKN. "Legacy" data are those collected prior to development of the I&M Program. By design, this perception of the CAKN data universe focuses on the monitoring mission.

Currently (12/15/2004), an amorphous mass of legacy data represents the bulk of the CAKN data family; however it is anticipated that monitoring data will eventually surpass all other categories in terms of volume. Since any data of sufficient vintage may be termed “legacy”, its relative abundance will persist. However, the intent is that greater portions of it will become less amorphous overtime as more rigorously managed data fall under that term.

Monitoring Data

The CAKN has identified 37 vital signs for long-term monitoring. Table 5.1 summarizes the types of data, frequency and timing of collection, and park where the data will be collected. The data represented in Table 5.1 comprise the primary target of the data management strategy and principles laid out in this plan. As noted in section IV.1, monitoring and supporting documents will be made available regionally via a secure server in the YUGA office. That portion of program output deemed fit for public use will be served via NPSpecies, NatureBib and NR-GIS out of WASO (see Appendix C).

Data for several of the vital sign parameters are collected and managed by entities other than the CAKN. These vital signs are marked with an asterisk in Table 5.1. Standard operating procedures for incorporating these data with the main body of CAKN products are included in the protocols for those vital signs.

Table 5.1. Summary of CAKN vital sign measures. Asterisks indicate data are managed outside the CAKN by specific protocol.

National Vital Signs Frame-work		Vital Sign Name	Park	Measures	Datatype	Frequency	Collection	Lab Data	Data Products
Level 1	Level 2								
Air and Climate	Air contaminants	Air quality*	DENA	pH, specific conductance, Ca, Mg, K, Na, NH4, NO3, Cl, SO4, and PO4, Belfort Recording Rain Gauge, aerosol mass, elemental and organic carbon, SO2, NO3, elements Na - Pb, H, N, and PM10 mass; IMPROVE monitoring network for visibility and fine particles	tabular	continuous	continuous		weekly and hourly summaries
	Weather and Climate	Climate*	CAKN	Temperature, precipitation, relative humidity	satellite download (ascii)	continuous	continuous		
		Snow pack*		Annual deposition	tabular	continuous	Sep. - May.		
Geology and Soils	Glacial features and processes	Glaciers	DENA WRST	Photo points of specified glaciers, benchmark measurement of glacier	tabular; imagery		Mar. – Apr.		
	Volcanic features and processes	Disturbance - volcanoes and tectonics*	CAKN	Frequency of volcanic activity, magnitude of volcanic activity, magnitude of tectonics			continuous		
		Permafrost		Geographic extent, boreholes	tabular, spatial		Jun. – Aug.		
Water	Surface water dynamics	Disturbance - Stream flood frequency and discharge*	CAKN	Timing of flood events, discharge rate of streams			Mar. – May.		
		River/stream flow		Annual hydrograph of selected rivers			continuous		
	Water chemistry	Water Quality		Water chemistry, water level of shallow lakes	tabular	2 per yr	Jun, Aug*	invertebrates, water samples	
	Aquatic macroinverts and algae	Macroinvertebrates			tabular	3 per yr	Jun, Aug*	invertebrates, water samples	

National Vital Signs Frame-work		Vital Sign Name	Park	Measures	Datatype	Frequency	Collection	Lab Data	Data Products
Level 1	Level 2								
Biological Integrity	Invasive/ Exotic plants	Disturbance - Exotic species	CAKN	Geographic extent of exotic plants			Jun. – Aug.		
	Insect pests	Insect Damage		Presence of insect damage on foliage	tabular	annual (air work), 5-10yr plot visits	Jul. – Aug.		
	Fishes	Freshwater fish		Population estimates of key species, geographic extent			Jun. – Aug.		
	Birds	Bald Eagles	WRST	Nesting territory occupancy, nesting success, mean brood size, overall population productivity, eggshell thickness, levels of contaminants	tabular	2 per yr, 5-8 yr contaminant sampling	mid May, late Jul.*	feather, egg, blood	
		Golden Eagles	DENA	Nesting territory occupancy, nesting success, mean brood size, overall population productivity, nesting territory fidelity, survival rates	tabular	2 per yr	late Apr., late Jul.*	DNA analysis, feather, egg, blood	
		Passerines	CAKN	Population trends of common species	tabular	annual	Jun. – Aug.		
		Peregrine Falcons	YUCH	Nesting territory occupancy, nesting success, mean brood size, overall population productivity, loads of organochlorine pesticides, mercury, eggshell thickness, breeding range habitat change	tabular	2 per yr	Jul. – Aug.*	feather, egg, blood	
		Ptarmigan	CAKN	Population density estimates, geographic extent	tabular	annual	Apr. – Jun.		
	Mammals	Arctic ground squirrels	DENA WRST	Ground squirrel distribution and abundance, habitat associations of ground squirrels, timing of hibernation periods	tabular, spatial	annual	Apr. – Aug.		
		Snowshoe hare	CAKN	Density of hare, geographic extent	tabular, video (possible)	annual	Jun. – Jul.		
		Small mammals		Density estimates of small mammals, geographic extent	tabular	annual	Jun. – Jul.		
		Caribou		Population estimate of herd size, geographic extent	tabular	2 per yr	Jun., Sep. - Oct.		

National Vital Signs Frame-work		Vital Sign Name	Park	Measures	Datatype	Frequency	Collection	Lab Data	Data Products
Level 1	Level 2								
				Population estimate, geographic extent	tabular	3yr	first snow : late Oct. in DENA; early-mid Nov. WRST,YUCH		
		Moose		Population estimate, geographic extent	tabular	annual	Jul. – Aug.		
		Sheep		Population estimate, geographic extent	tabular	annual	Jul. – Aug.		
		Wolves		Population estimate, geographic extent	tabular, email	3 per yr	Nov., Feb., Jun.*	genetic sampling/analysis, blood/tissue	
		Brown bear		Density estimate of brown bears, geographic extent	tabular	once every 2 yrs or annually	Mar. – May		
	Vegetation communities	Vegetation structure and composition		Absolute and relative abundance of growth-form classes, abundance and composition of dominant species, distribution and abundance of discrete vegetation types, species richness, species composition, basal area of tree species, depth of active layer, species diversity, species density, fuel load, type size and position of fuels, depth of duff layer, depth of litter layer	tabular, digital photo	annual	Jun. – Aug.*	soils, tree cores	
	Terrestrial communities	Subarctic Steppe			tabular, spatial		Jul.		
Human use	Point-source human effects	Human populations*	CAKN	Population trends in settlements adjacent to network parks from U.S. Census Bureau data	tabular	US Census-10yrs, AK data- <10yrs			
	Consumptive use	Consumptive uses of National Park natural resources*		Fish and wildlife harvest (numbers and locations), marine derived nutrients, number of trees harvested for fire and house logs	tabular	occurrence-based (hunting, seasonal, etc.)			
	Visitor usage	Human Presence/Use*			tabular	generally summer			

National Vital Signs Frame-work		Vital Sign Name	Park	Measures	Datatype	Frequency	Collection	Lab Data	Data Products
Level 1	Level 2								
		Trails			tabular	Jun-Aug			
Ecosystem Pattern and Processes	Fire and fuel dynamics	Disturbance - Fire occurrence and extent	CAKN	Long-term trend of fire freq, average fire size, average burn severity, total area affected by fire	tabular, image, spatial	occurrence-based	Jun. – Aug.		
	Land cover / Land use	Landcover		Percent of network in specified land classes, vegetative cover	imagery	annual	Mar., Oct.		
	Soundscapes	Sound		Biophonics of backcountry	tabular; recordings	May – Oct.			
	Productivity	Forage quantity/quality		Nitrogen content of Salix sp. Leaves, geographic extent of Salix sp.	tabular		Jun. – Aug.*	pellet analysis (WSU, UAF labs); foraged plant analysis	
		Plant phenology		Snow-free data, date of onset of greenness, date of maximum greenness, date of senescence of greenness, snow-cover date			May – Sep.		

Short-term Projects

Pilot projects, contracts for targeted data collection, and other short-term projects are anticipated to help mold or redirect the monitoring program over time. In development of the monitoring program, the CAKN has implemented several pilot projects designed to explore different monitoring possibilities. Currently (12/15/2004), the short-term data represents the latest data acquisitions a large portion of which will likely join the ranks of primary monitoring data once monitoring protocols are finalized.

Inventory

The twelve basic resource inventories summarized in Appendix A comprise the bulk of the inventory data. A majority of these are managed outside the CAKN monitoring program as indicated in Appendix A. However, biological inventories of birds (1998), small mammals, fish and vascular plants were conducted between 2000 and 2004 in network parks and are directly managed by the CAKN. Final reports for these inventory projects are expected in 2005. Inventory data and reports are (and will be) catalogued in the nationally maintained databases NPSpecies (species counts), NatureBib (reports) and NR-GIS (data and metadata). These end-products will also be available via the CAKN website and data-server.

Prototype (DENA LTEM)

Between 1991 and 2002, the Denali LTEM program collected data culminating in nineteen distinct datasets within the disciplines of glaciers, snowfall, weather, soils, vegetation, aquatic systems, aquatic invertebrates, wolf/prey interactions, small mammals, eagles/gyrfalcons, and passerines. A final report on the Denali LTEM program data was submitted in 2003 (*www1.nature.nps.gov/im/units/cakn/Documents/DENA_DataMgtReport.pdf*). These data help form a basis for monitoring efforts in the CAKN.

Legacy

The term “legacy” may generally be applied to data collected prior to development of the I&M Program. Data mining exercises within the CAKN have documented over 300 datasets that have been generated by the CAKN parks between 1970 and 2003. The process of identifying, collecting, evaluating and developing older data sets related to the CAKN’s mission is expected to continue over time. Section V.2 addresses the acquisition of data and information via data mining.

V.2 Acquisition through Data Discovery/Data Mining

Data discovery or data mining is the process of searching for existing data/information that may be useful to the I&M Program mission and that is related to the natural resources of the network parks.

A large percentage of data discovery occurs at the onset of new projects or during the development of new protocols. The process involves reviewing many different sources for varying types of information. Many of the following data sources are accessible via the internet but some require visiting local research or academic institutions, museums or local parks to conduct the searches.

Bibliographic/Literature

- National NPS Databases (e.g. NatureBib)
- Online literature databases (e.g. ARLIS (see Chapter IV), First Search, Biosis)
- Local document library (e.g. BLM, USFWS libraries)
- Library catalogs (e.g. academic or research institutions)
- Park archives

Geographic Data

- Regional GIS Specialists
- Park GIS Specialists
- Federal and State Geographic data clearinghouses

Biological/Natural Resources Data

- NPSpecies
- Voucher collections (museums, parks, universities)
- Network Parks

All information collected during data discovery process is maintained at the CAKN-member park from which the discovery was made either electronically or in hard copy format (depending on how it was collected). A catalog of all information is maintained at the CAKN office (NPS YUGA office) and copied to the NPS NR-GIS metadata and data store (see Appendix C). Any geographic datasets collected during this process should be accompanied by FGDC compliant metadata. Chapter IX details dissemination mechanisms and repositories for CAKN data and information.

If legacy data are collected in a digital format, the information should be converted to current file formats compatible with the current software standards. In the past, hard copy materials were maintained as such but in the future, the CAKN intends to scan hardcopy references and materials, saving them as .pdf files, in order to create a digital library.

STRATEGY
The CAKN intends to scan hardcopy references and materials, saving them as .pdf files, in order to create a digital library of legacy data.

Data discovery is an integral part of project development but data discovery efforts are not limited to project development needs. It is an on-going process requiring regular data searches and visits to Network parks in order to ensure that the CAKN I&M Program maintains as much relevant material pertaining to the parks as possible. Encouraging data sharing with Network parks will assist in this process and may alleviate the need for regular searches of park records.

Data discovery efforts are not limited to project development needs. It is an on-going process requiring regular data searches and visits to Network parks.

V.3 Acquisition of Monitoring Data (Timing)

The timing for data acquisition and other critical data management steps for the monitoring vital signs chosen by the CAKN is summarized in Table 5.2. Data for each of these vital signs will enter and flow through the system illustrated in Figure 4.4 on a timeline shown in Table 5.2.

Vital signs followed by an asterisk in Table 5.2 represent data initially collected and managed by entities other than the CAKN. Standard operating procedures for incorporating these data with the main body of CAKN products are included in the protocols for those vital signs.

Table 5.2. CAKN vital sign data processing timing and products. Each vital sign will at a minimum have a report and GIS layer as a product. A "†" symbol next to the collection timing indicates additional laboratory analysis for the vital sign.

Lvl 1	Vital Sign Name	Collection	Entry/ Checked	Analysis report	Integrated
Air and Climate	Air quality*	continuous			
	Climate*	continuous			
	Snow pack*	Sep. - May.	Jul.	Jul.	Aug.
Geology and Soils	Glaciers	Mar. – Apr.	Jun.	Jan.	Feb.
	Disturbance - volcanoes and tectonics*	continuous			
	Permafrost	Jun. – Aug.	Oct.	Apr.	May
Water	Disturbance - Stream flood frequency and discharge*	Mar. – May.	Jul.	Jan.	Feb.
	River/stream flow	continuous			
	Water Quality	Jun, Aug†	Sep.	Apr.	May
	Macroinvertebrates	Jun, Aug†	Sep.	Apr.	May
Biological Integrity	Disturbance - Exotic species	Jun. – Aug.	Oct.	Apr.	May
	Insect Damage	Jul. – Aug.	Oct.	May	Jun.
	Freshwater fish	Jun. – Aug.	Oct.	Apr.	May
	Bald Eagles	mid May, late Jul.†	Sep.	Mar.	Apr.
	Golden Eagles	Apr., late Jul.†	Sep.	Feb.	Mar.
	Passerines	Jun. – Aug.	Oct.	Apr.	May
	Peregrine Falcons	Jul. – Aug.†	Oct.	Apr.	May
	Ptarmigan	Apr. – Jun.	Aug.	Feb.	Mar.
	Arctic ground squirrels	Apr. – Aug.	Oct.	Feb.	Mar.
	Snowshoe hare	Jun. – Jul.	Sep.	Apr.	May
	Small mammals	Jun. – Jul.	Sep.	Apr.	May
	Caribou	Jun., Sep. -Oct.	Dec.	Apr	May
	Moose	first snow (late Oct. in DENA; early-mid Nov. WRST, YUCH)	Jan.	Aug.	Sep.
	Sheep	Jul. – Aug.	Oct.	May	Jun.
	Wolves	Nov., Feb., Jun.†	Jan, Apr, Aug	Sep.	Oct.
	Brown bear	Mar. – May	Jul.	Jan.	Feb.
	Vegetation structure and composition	Jun. – Aug.†	Oct.	Apr.	May
	Subarctic Steppe communities	Jul.	Sep.	May	Jun.
Human Use	Human populations*	US Census-10yrs, AK data-<10yrs			
	Consumptive uses of National Park natural resources*	occurrence-based (hunting, seasonal, etc.)			
	Human Presence/Use*	generally summer			
	Trails	Jun. – Aug.	Oct.	Apr.	May
Ecosystem Pattern and Processes	Disturbance - Fire occurrence and extent*	Jun. – Aug.	Oct.	Apr.	May
	Landcover	Mar., Oct.	Dec.		
	Sound	May – Oct.	Dec.	Mar.	Apr.
	Forage quantity/quality	Jun. – Aug.†	Oct.	Apr.	May
	Plant phenology	May – Sep.	Nov.	Mar.	Apr.

V.4 Initial Processing after Acquisition

Chapter IV presented the general data life cycle and flow in Figure 4.4. Data dissemination for the CAKN is presented in Chapter IX. While the standard operating procedures for each collection effort offer specifics, in general raw data collected by the Network are archived immediately following completion of field work. This usually entails photocopying field sheets, downloading GPS coordinates and possibly saving data collected on field-based computers to local drives. Each vital sign protocol includes an SOP that describes the immediate handling of raw data upon return from field work.

Raw data are archived immediately following completion of field work. Each vital sign protocol includes an SOP that describes the immediate handling of raw data upon return from field work.

Working Databases

As indicated in Figure 4.4, a working database is used to enter, validate and verify data prior to metadata generation and upload to the primary CAKN data server. Each monitoring vital sign uses a working database developed in MSAccess to perform this initial data processing. These databases, like the primary client/server database, are built in compliance with the I&M Program Natural Resource Database Template (NRDT, *science.nature.nps.gov/im/apps/template*). Each database is tailored to the needs of the project under the supervision of the Network data manager (see Chapter III for roles and responsibilities).

Field Specimen and Samples

Protocols for each Vital Sign monitoring project specify the acquisition and initial processing of material collected in the field. The protocol for each vital sign requiring laboratory analysis dictates the transfer of specimen and/or samples and subsequent acquisition of analytical results. The Network imposes the following strictures regarding laboratory analysis for vital sign monitoring:

- As part of data collection and entry duties (see Chapter III), the project leader and data manager will ensure that analytical data match the project data design and formats.
- The data manager and project leader must build into the project data design a mechanism for relating instances of field work (data collection events), raw data, and laboratory results. This will be done via sample and specimen labels that can be matched with field work event identification numbers.

STRATEGY
The project leader and data manager ensure that:
- Analytical data match project data design and formats
- Analytical data can be matched with field events which are in turn related to other data collected in the field

Chapter VI presents additional details regarding quality control and assurance for CAKN data. Chapter IX identifies the repositories the CAKN uses for field specimen and samples.

Specialized Data Acquisition

In some cases, raw data requires specialized processing before actual analysis can proceed. Examples of such data include photography or satellite imagery requiring interpretation or filtering before analysis for vital sign measures may begin. In these cases, like in the case of laboratory data, the project protocol specifies the procedures used to acquire and process data in preparation for analysis. Here again, it is the responsibility of the project leader and network data manger to ensure that data fitting this rubric match the project's overall data design and can be related to any additional parameters collected by the project.

V.5 Changes to Data Collection Procedures/Protocols

As indicated in the generalized project process (Figure 2.1), long-term monitoring project data are evaluated to determine if project methods are meeting stated goals. Ideally, problems with data collection procedures are identified and corrected during the design and testing stage of a project. Also during this stage, potential problems are ideally identified and contingency procedures established prior to data collection.

Protocol changes, however, are inevitable. Significant changes to the protocols must be approved by the project leader, network coordinator and the data manager. The network coordinator must evaluate the proposed changes and determine if additional peer review is required before approving.

Chapter VI specifies that as part of general quality assurance, CAKN database designs will include record-level tracking that indicates the protocol version under which data were collected and processed.

VI. Quality Assurance (QA) and Quality Control (QC)

We must have confidence in the data we use. Our analyses–to detect trends or patterns in ecosystem processes–require data of documented quality that minimize error and bias. Data of inconsistent or poor quality can result in loss of sensitivity and incorrect interpretations and conclusions. The CAKN monitoring program will bring together analysis from an array of widely-varying projects aimed at detecting changes in the environment over time. The inherent complexity in this endeavor demands stringent and consistent quality assurance and control measures be applied throughout the program.

Quality assurance involves planning quality into the data while quality control consists of monitoring the system or appraising the product after the product is developed. Data management for the CAKN Inventory and Monitoring program must ensure that our projects produce and maintain data of the highest possible quality. The Network will establish and document protocols for the identification and reduction of error at all stages in the data lifecycle. These stages include project planning and database design, data collection, data entry, verification and validation (certification), documentation (including data quality and sensitivity review), processing, and archiving (Figure 6.1). The final stage in the data life course is dissemination/integration. Detailed QA/QC procedures for these stages are included in the protocols for each project initiated by the CAKN. This chapter of the data management plan presents more broadly based procedures and policy that govern specific operations within a CAKN project. Figure 6.2 illustrates selected QA/QC procedures in context of the amount of control necessary and data confidence.

STRATEGY
The Network will establish and document protocols for the identification and reduction of error at all stages in the data lifecycle.

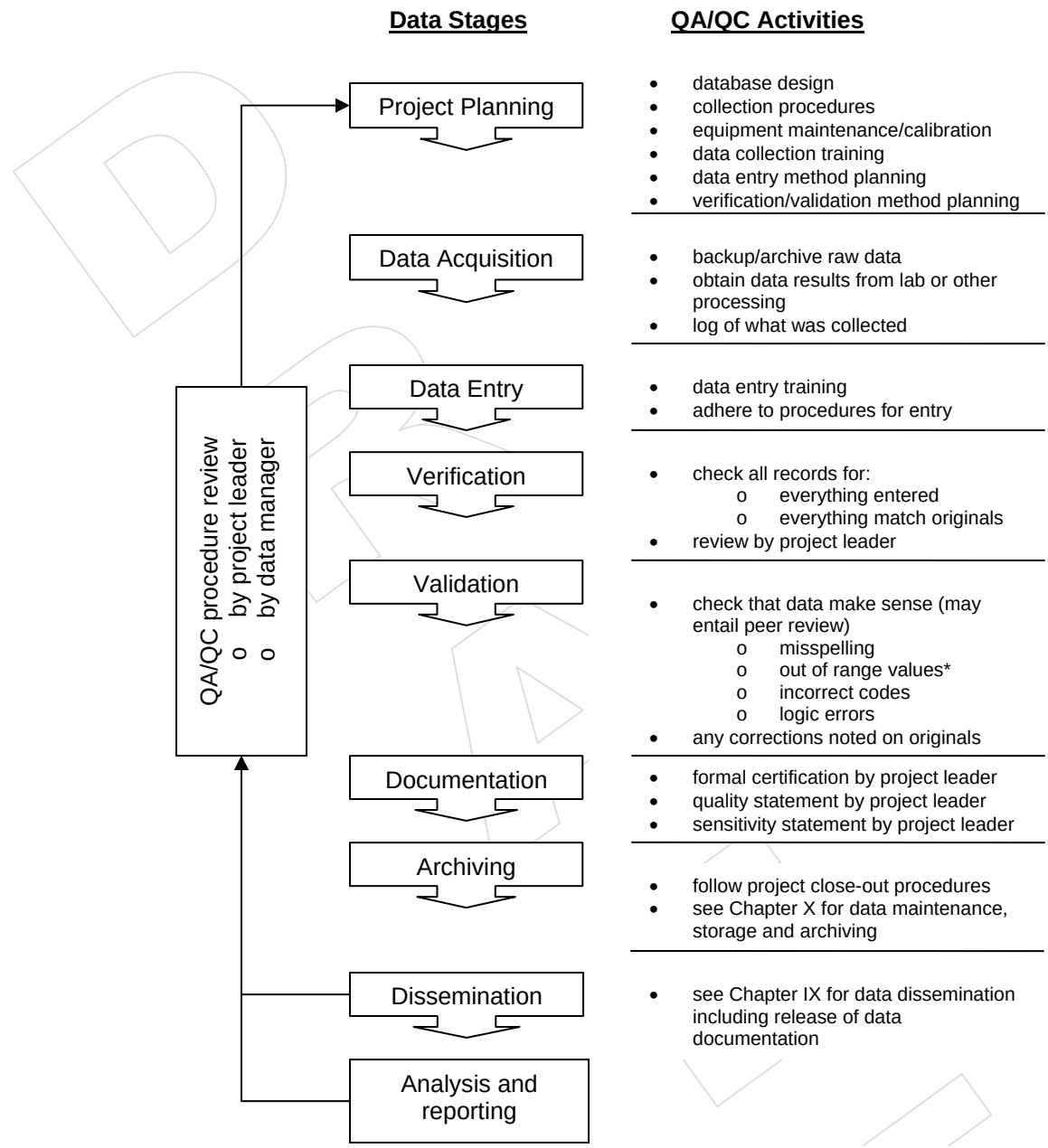
The treatment of QA/QC in this plan can not be properly done without specifying certain roles and responsibilities native to such procedures. The duties outlined in this chapter are inherited from those listed in Chapter III and conform to their letter and spirit. These duties should be followed in light of those given in Chapter III.

A Word on Data Quality Expectations

Although a data set containing no errors would be ideal, the cost of attaining 95%-100% accuracy may outweigh the benefit. Therefore, we consider at least two factors when setting data quality expectations:

- frequency of incorrect data fields or records
- significance of error within a data field

We are more likely to detect an error when we work with clearly documented data sets and understand what a “significant” error is within *that* data set. The significance of an error can vary with data sets and depends on where it occurs. For example, a two-digit number off by one decimal place is a significant error. A six-digit number, with the sixth digit off by one decimal place, is not a significant error. But one incorrect digit in a six-digit species number could indicate a different species. That is a significant error.



* CAUTION! Care must be exercised when culling "out-of-bounds" data. See text for discussion.

Figure 6.1. General course of data and associated QA/QC procedures. Quality control with regards to data analysis is specific to each project and addressed in appropriate standard operating procedures.

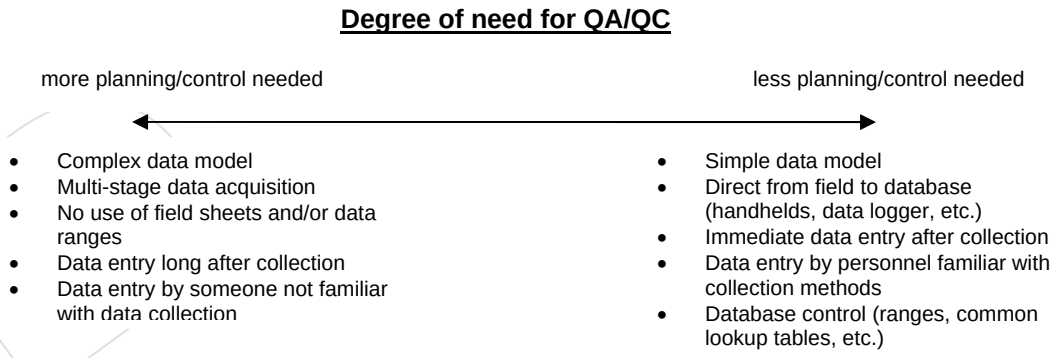


Figure 6.2. Some common data management elements affecting degree of need for QA/QC. Planning and training for data collection (QA) and entry is always a premium.

VI.1 National Park Service Mandate for Quality

NPS Director's Order #11B: "Ensuring Quality of Information Disseminated by the National Park Service," issued in 2002, promotes information and data quality. It defines *'quality'* as incorporating three key components–*objectivity*, *utility*, and *integrity*.

Objectivity consists of: 1) *presentation,* which focuses on whether disseminated information is being presented in an accurate, clear, complete, and unbiased manner within a proper context, and 2) *substance,* which focuses on ensuring accurate, usable, and reliable information.

Utility refers to the usefulness of the information to its intended users, from the perspectives of both the Network and the general public.

Integrity refers to the *soundness* of the data or the confidence one has in the data. Integrity is integrally related to objectivity; however it is possible to have subjective data of high integrity. The integrity of data is also related to data security, e.g., protection from unauthorized access or revision to ensure that the information is not compromised through corruption or falsification.

Order #11B also specifies that information must be based on reliable data sources, which are accurate, timely, and representative of the most current information available. These standards apply not only to NPS-generated information, but also to information provided by other parties to the NPS if the NPS disseminates or relies upon this information.

High quality data and information are mandated by directives and orders, and they are vital to the credibility and success of the I&M program. According to Abby Miller (2001) of the Natural Resource Stewardship and Science Division, "data need to meet national-level quality standards and need to be accessible to be used

These standards apply not only to NPS-generated information, but also to information provided by other parties to the NPS if the NPS disseminates or relies upon this information.

for wise and defensible decision-making at all levels. Data need to be able to be shared and aggregated with data from other parks and from adjacent lands to support landscape-level and national planning and decision-making."

VI.2 Quality Assurance and Control Duties

Everyone involved with the handling of project data plays a part in producing and maintaining high quality data. While Chapter III specifies data management roles and responsibilities, selected QA/QC duties are emphasized here.

Project managers must:

- be aware of quality protocols and convey their importance to technicians and field crews
- ensure compliance with the protocols
- plan for and ensure proper execution of data verification and validation
- review all final reports and information products

Technicians must follow established protocols for data collection, data entry, and verification established in the project SOPs.

The data manager is responsible for:

- developing Network-wide protocols and SOPs to ensure data quality
- making project managers, technicians, etc., aware of the established procedures and enforcing adherence to them
- evaluating the quality of all data and information against NPS standards before dissemination outside the network
- performing periodic data audits and quality control checks to monitor and improve data quality control operations

VI.3. Data Quality Goals and Objectives

We must ensure that a project produces data of the right type, quality, and quantity to meet project objectives. Quality criteria should be set at a level proportionate to project-specific objectives and these criteria should indicate the level of quality acceptable for resulting data products. To reiterate, project subjects and goals will drive data quality needs and control the kinds of analysis and summarization that may be defensibly applied.

Project subjects and goals will drive data quality needs and control the kinds of analysis and summarization that may be defensibly applied.

STRATEGY
The CAKN will develop a comprehensive set of SOPs and tools for quality assurance and control in field procedures, data entry/validation/verification and data use (browsing, sub setting, downloading, analysis, etc.).

The most effective mechanism for ensuring that a project produces data of the right type, quality, and quantity is to provide procedures and guidelines to assist the researcher in accurate data collection, entry, and validation. As part of data management operations, the CAKN will develop a comprehensive set of SOPs and tools for quality assurance and control in field procedures, data entry/validation/verification and data use (browsing, sub setting, downloading, analysis, etc.). This will likely include development of new field-computer-based methods for acquiring data as well as custom programming of

database interfaces to both protect the data and control how the data are viewed or summarized.

VI.4. General Operations

File Naming Standards

Because of the scope and quantity of files being consolidated into one place, it is critical that computer files be given specific and unique names that will uniquely identify them indefinitely. The basic file structure for CAKN digital information is outlined in Chapter II digital information.

The following conventions apply to all CAKN file names:

- No spaces or special characters within the name
- Use date for version control, as described below
- Use underscore as delimiters
- Names shall be 30 characters or less

File names for final products will begin with “CAKN” and contain a brief but clear explanation of the file content followed by an eight digit date and the extension. Each of these file name elements (excepting the extension) is preceded by an underbar (“_”). Additional underbars may be added for clarity. As an example “CAKN_draftData_Management_Plan_11132004.doc” indicates a draft version of the data management plan dated Nov. 13, 2004.

Version Control

Prior to any major changes to a file, a copy of the file with the appropriate version control (in this case the file name is the version control) is made. As indicated above, files are stored with the appropriate eight digit date which serves as version control. This allows the tracking of changes over time and facilitates collaboration between multiple personnel working on common files. With proper controls and communication, versioning ensures that only the most current version is used in any analysis. Using the date, formatted as MMDDYYYY, provides a logical version control.

Laboratory Data

Several of the CAKN vital signs involve laboratory analysis conducted via contract with established laboratories. All data for a given vital sign, including lab results, will be housed in a distinct relational database accessed via custom applications built in MS Access. Laboratories that will be entering analysis results for a given vital sign will be supplied with a copy of the application so that data may be entered in the manner and format matching that of the rest of the data for a given monitoring parameter. While most professional laboratories exercise their own QA/QC procedures, results received by a project leader are subject to the same QA/QC measures exacted on other project data.

VI.5. Project Planning and Data Design (Quality Assurance)

The methods and array of information a given project chooses to use and gather are integral to data quality. The CAKN holds as policy that techniques and procedures chosen for a project will maximize data quality. This will be achieved by stipulating that:

- Common lookup tables for values of parameters recorded in the field by more than one project (such as common weather metrics)

- Each project has SOPs for core data management (collection, entry, verification, etc.)
- Each project use, at a minimum, standard field sheets for data collection (see Section VI.6)
- Database data entry computer applications resemble field sheets
- Automated error checking features be included in database applications (see Sections VI.6, VI.7 and VI.9)
- Database application design will maximize the use of auto-fill, auto-correct, value range limits, pick lists, and other constraints specific to projects
- Database applications will include a means to track errors reported on the data after dissemination (see Chapter IX).
- Database maintenance logs will be maintained for each CAKN database and housed in association with database files.

More on Database Design - Record-level Tracking

As a standard part of database design, the CAKN will build into database tables, fields that track at the record-level who entered the data, precise entry time and the protocol version under which the data were collected. The benefits to overall data integrity outweigh any inconveniences this "overhead" data may cause due to factors such as increased database file size.

STRATEGY
As a standard part of database design, the CAKN will build into database tables, fields that track at the record-level who entered the data, precise entry time and the protocol version under which the data were collected.

More on Lookup Tables

As noted in Section II.1, the CAKN will utilize to the fullest extent possible common lookup tables for variables recorded by multiple projects. Examples include weather variables (such as precipitation intensity, wind speed, etc.), standard equipment and settings (e.g., GPS models and datums) and possible field personnel. Section VI.9 addresses database programming used for data validation.

More on Project SOPs

Each vital sign protocol will include SOPs that address core data management practices with quality control in mind. These may include:

- Field crew training (addressing both data collection and entry)
- Standardized data sheets
- Use of handheld computers
- Equipment maintenance and calibration
- Procedures for handling data (including specimen) in the field
- Data backup, entry, verification and validation

VI.6. Data Collection

Chapter V addressed basic acquisition and initial handling of data. Attention to detail during data collection is crucial to overall data quality. The CAKN adopts the following precepts regarding data collection that affect data quality. Changes to data collection protocols are addressed in Section V.5.

- At a minimum, data will be collected on formatted, project-specific data sheets that reflect the overall data design for the project and maximize limitations on

values that may be recorded for different parameters. Sheets will be designed to minimize the amount of writing necessary to effectively record observations. Data sheets will be printed on a stock appropriate to field conditions (generally "write-in-the-rain" paper).

- Any project using field equipment will include an SOP for calibration (including the timing of calibration) and maintenance. Such SOPs will specify establishment of an equipment maintenance log.
- Edits on field sheets while in the field will be done by drawing a single line through the information to be changed and adding any replacement information in clear writing next to the original entry.
- Completed field forms will be proofed for errors each day in the field.
- Wherever possible and appropriate, data loggers or field-based computers will be used to collect data. The use of handheld computers requires separate SOPs to describe their use and will include direction for daily review and back up of digital data.

More on Field Sheets

Standardized data sheets that identify the pieces of information to be recorded and forms that reflect the design of the computer data entry interface will help ensure that all relevant information is recorded and subsequent data entry errors are minimized. Data sheets should contain as much basic preprinted project information as possible and sufficient space for recording relevant metadata such as date, collectors, weather conditions, etc. They should clearly specify all required information, using examples where needed to ensure that the proper data are recorded. Data recorders should adhere to the following guidelines:

- All information added to the data sheet must be printed and clearly legible.
- If alterations to the information are necessary, the original information should be crossed out with a single line and the new information written next to the original entry. Information should never be erased and old information should not be overwritten.
- Upon return from the field, copies of all original data sheets should be made and checked for legibility and completeness (i.e., no data cut off at the edges). The copies of the data sheets will be stored as specified in the protocol SOP, and the original data sheets will be used for data entry.

VI.7. Data Entry or Import

Data entry is a separate operation from data validation and care must be taken to not impose validation (beyond that automatically imposed by programming rules in a database) during data entry.

'Data entry' is the initial set of operations where we transfer raw data into a computerized form linked to database tables. When data are gathered or stored digitally in the field (e.g., on a data logger), data entry consists of the transfer of data (downloading) to a file in an office computer where they can be further manipulated. The goal of data entry is to transcribe field observations into a computer database with 100% accuracy. That is, exactly what was recorded in the field should be entered into the database. Subsequent data verification (see Section VI.8) is conducted to ensure that raw data matches entered data. Following verification, data validation may result in changes *to the entered data.* Data entry is a separate operation from data validation and care must be taken to not impose validation

(beyond that automatically imposed by programming rules in a database; see Section VI.9) during data entry.

The CAKN adopts the following precepts affecting data quality with regards to data entry:

- Data will be entered as soon as reasonably possible after collection.
- Data entry will be done by someone familiar with data collection. The project leader (with assistance from the data manager if needed) must ensure that data entry staff is familiar with the database software, database structure, and any standard codes for data entry used by the Network. At minimum, data entry technicians should know how to open a data entry form, create a new record, and exit the database properly. They must learn how to commit both a 'field' entry and a "complete record" entry and to correct mistakes made while typing.
- If possible, data will be entered by two qualified persons; one to read the observations and one to enter.
- Data will be entered into pre-designed database applications that resemble field sheets and maximize error control (see Sections VI.6 and VI.9). Data will not be entered into spread sheets.
- To the extent possible, data entry will be automated. This may simply entail downloading of data from field-based computers but may include the application of new technology to allow for machine-driven data entry.
- The CAKN will maximize the use of database programming to control data entry (see Sections VI.5, VI.6 and VI.9). In general, this will be achieved via the use of lookup tables (see Section VI.5) but may be accomplished by field-type design in a data base (such as "yes/no" field-types).

VI.8. Data Verification (Quality Control Part 1)

We appraise data quality by applying verification and validation procedures as part of the quality control process. These procedures are more successful when preceded by effective quality assurance practices (planning). *Data verification* checks that the digitized data match the source data, while *data validation* checks that the data make sense. It is essential that we validate all data as truthful and do not misrepresent the circumstances and limitations of their collection. Failure to follow SOPs for data entry, validation, and verification will render a data set suspect. Although data entry and data verification can be handled by personnel who are less familiar with the data, validation requires in-depth knowledge about the data.

The CAKN adopts the following precepts regarding data verification:

- Options for verification methods are given below. Project leaders are responsible for specifying in the project protocol one or more of these methods. At the discretion of the project leader, additional verification methods may be applied.
- The project leader is responsible for proper execution of data verification (see Section VI.2 and Chapter III).
- Data verification is carried out by staff sanctioned by the project leader who are ideally familiar with data collection and entry.
- 100% of records will be verified against original source data.
- 10% of records will be reviewed after initial verification by the project manager. If errors are found, the entire data set verified again.

- A record of each dataset's verification process including number of verification iterations and results will be prepared by the project leader as part of formal metadata generation (see Section VI.10 and Chapter VII for more details).

Methods for Data Verification

Each of the following methods has a direct correlation between effectiveness and effort. The methods that eliminate the most errors can be very time consuming while the simplest and cheapest methods will not be as efficient at detecting errors.

1) *Visual review at data entry.* The data entry technician verifies each record after input. Recorded values in the database are compared with the original values from the hard copy. Identified errors are immediately corrected. This method is the least complicated since it requires no additional personnel or software. Its reliability depends entirely upon the person keying data and thus, is probably the least reliable data verification method.

2) *Visual review after data entry.* All records are printed upon completion of data entry. The values on the printout are compared with the original values from the hard copy. Errors are marked and corrected in a timely manner. When one technician performs this review, the method's reliability increases if someone other than the person keying data performs the review. Alternatively, two technicians can perform this review. One technician reads the original data sheets (the reader), and the second reads the same data on the printout (the checker).

3) *Duplicate data entry.* The data entry technician completes all data entry, as normal. Random records are selected (every **n**th record) and entered into an empty replica of the permanent database, preferably by someone other than the person keying the permanent data. A database query is then used to automatically compare the duplicate records from the two data sets and report any mismatches. Disparities are manually reviewed and correction applied if necessary. This method adds the overhead of retyping the selected records, as well as the creation of a comparison query. However, it becomes increasingly successful as the value of **n** decreases. Professional data entry services frequently use this method.

Supplementary Methods

- Calculate simple summary statistics with statistical software to identify duplicate or omitted records. For example, we can view the number of known constant elements, such as the number of sampling sites, plots per site, or dates per sample. We can pose the same question in different ways; differences in the answer provide clues to errors. The more checks we devise to test the completeness of the data, the greater our confidence that we have completely verified the data.

VI.9. Data Validation (Quality Control Part 2)

Validation is the process of reviewing computerized data for range and logic errors and may accompany data verification *only* if the operator has comprehensive knowledge of the data and subject. More often, validation is a separate operation carried out *after*

verification by a project specialist who can identify generic and specific errors in particular data types.

General step-by-step instructions are not possible for data validation because each data set has unique measurement ranges, sampling precision, and accuracy. Nevertheless, validation is a critically important step in the certification of the data and a required component of any CAKN project protocol (see Section VI.5). Invalid data commonly consist of slightly misspelled species names or site codes, the wrong date, or out-of-range errors in parameters with well defined limits (e.g., pH). But more interesting and often puzzling errors are detected as unreasonable metrics (e.g., stream temperature of 70°C) or impossible associations (e.g., a tree 2 feet in diameter and only 3 feet high). We call these types of erroneous data *logic errors* because using them produces illogical (and incorrect) results. The discovery of logic errors has direct, positive consequences for data quality and provides important feedback to the methods and data forms used in the field. Histograms, line plots, and basic statistics can reveal possible logic and range errors.

The CAKN adopts the following precepts regarding data validation:

- Each CAKN project protocol will address a process for data validation that includes at least one of the methods outlined in the Network data management plan. That process will adhere to the precepts given in this plan.
- Corrections or deletions as a result of data validation require notations in the original paper field records and in any copies made for data entry about how and why the data were changed.
- Modifications of the field data should be clear and concise while preserving the original data entries or notes (i.e., no erasing!).
- Validation efforts should also include a check for the completeness of a data set since field sheets or other sources of data could easily be overlooked.
- The CAKN will maximize the use of automated routines and/or data summary/visualization such as histograms, line plots, and basic statistics to reveal possible logic and range errors.

Methods for Data Validation

The following general methods can be used to validate data. Specific procedures for data validation depend upon the vital sign being monitored (project subject) and will be included in the project protocols.

1) *Data entry application programming.* Certain components of data validation are built into data entry forms (see Section VI.7). This method is essentially part of database design and is noted in Sections VI.5, VI.6 and VI.7). Not all fields, however, have appropriate ranges that are known in advance, so knowledge of what are reasonable data and a separate, interactive validation stage are important.

   Caution must be exercised when using lookup tables to constrain variable values. Values occurring outside the range set by a lookup table (established during database design) may not always be invalid. As part of data validation procedures, the project leader is responsible for correct use of lookup tables or other automated value range control.

2) *Outlier Detection.* According to Edwards (2000), "the term outlier is not (and should not be) formally defined. An outlier is simply an unusually extreme value for a variable, given the statistical model in use." Any data set will undoubtedly contain some extreme values, so the meaning of 'unusually extreme' is subjective. The challenge in detecting outliers is in deciding how unusual a value must be before it can (with confidence) be considered 'unusually' unusual.

   Data quality assurance procedures should not try to **eliminate** outliers. Extreme values naturally occur in many ecological phenomena; eliminating these values simply because they are extreme is equivalent to pretending the phenomenon is 'well-behaved' when it is not. Eliminating data contamination is a better way to explain this quality assurance goal. If contamination is not detected during data collection, it is usually only detected later if an outlying data value results. When we detect an outlier, we should try to determine if some contamination is responsible.

   We can use database, graphic, and statistical tools for ad-hoc queries and displays of the data to detect outliers. Some of these outlying values may appear unusual but prove to be quite valid after confirmation. Noting correct but unusual values in documentation of the data set saves other users from checking the same unusual values.

3) *Other exploratory data analyses.* Palmer and Landis (2002) suggest that in some cases, calculations for assessments of precision, bias, representativeness, completeness, and comparability may be applicable and that for certain types of measurements, evaluation of a detection limit may also be warranted (the authors provide examples of procedures that may be applicable). Normal probability plots, Grubb's test, and simple and multiple linear regression techniques may also be used (Edwards, 2000; the author provides SAS and Splus code for constructing normal probability plots and examples of output showing normal and non-normal distributions).

VI.10. Data Quality Review and Communication

The National Park Service requires QA/QC review prior to communicating/disseminating data and information. Only data and information that adhere to NPS quality standards may be released (see Section VI.1). Data and Information disseminated to the public must be approved by the appropriate reviewing officials and programs. Documentation of the QA/QC standards used in producing the information and that substantiate the quality of the information must be formally certified and distributed with related data and information. Also, mechanisms must be in place for receiving and addressing comments/complaints pertaining to the quality of data. Chapter IX addresses the dissemination of CAKN data and information.

Monitoring Conformance to Plans and Standards

As part of the close out and evaluation stage of each CAKN project, QA/QC procedures will be reviewed by the project leader and any recommendations for change made in the project annual report. Additionally, the CAKN data manager will review the QA/QC procedures included in the data management plan and revise as needed. While this review will be part of the regularly scheduled plan review (see Section I.5 and Appendix B), QA/QC procedures are subject to revision as needed by the Network data manager.

The data manager will conduct periodic “spot checks” of random CAKN monitoring projects to ensure compliance with data management plan and project protocol QA/QC procedures. The data manager will track and facilitate the correction of any deficiencies found during this process. The data manager will submit a report of findings to the project leader and the network coordinator within a month of completing any review of QA/QC procedures. The project manager and Network data manager are responsible for ensuring that non-conformities in data management practices are corrected.

STRATEGY
The Network data manager will conduct periodic “spot checks” of random CAKN monitoring projects to ensure compliance with data management plan and project protocol QA/QC procedures. The data manager will submit a report of findings within a month of completing any review of QA/QC procedures.

Periodic checks by the data manager to see if network staff are adhering to the data quality procedures established in the Data Management Plan and protocols SOPS may include verification of the following:

- Data collection and reporting requirements are being met
- Data collection and reporting procedures are being followed
- Verification and validation procedures are being followed
- Data file structures and maintenance are clear, accurate and according to plan
- Revision control of program documents and field sheets are adequate
- Calibration and maintenance procedures are being followed
- Seasonal and temporary staff have been trained in data management practice
- Metadata collection and construction for the program proceeds in a timely manner
- Data are being archived and catalogued appropriately for long term storage

Documenting and Communicating Quality

The final step in the QA/QC for a given dataset is the preparation of summary documentation that assesses the overall data quality. A statement of data quality will be composed by the project leader and incorporated into formal metadata as well as the CAKN primary data repository. Metadata for each dataset/database will also provide information on the specific QA/QC procedures applied and the results of the review. Typically, data quality information will be conveyed as part of FGDC-compliant metadata (see Chapter VII for data documentation). Metadata and data will be available via both the CAKN website and the NPS NR-GIS natural resources data store (see Chapter IX for dissemination information).

VII. Data Documentation

The information needed to understand and use data is embedded in project documents such as study plans, reports, and protocols. Data design documents such as database descriptions and field sampling protocols are often critical to effective analysis of project data. Project tracking applications used by managers may also contain indispensable information about project data. Further, standard data handling such as backup, entry, verification and validation may produce information that affects the way in which the data may be used.

Formal, standard metadata serves as a unifying document for all these sources of information about project data (Figure 7.1). While metadata are addressed in context throughout this plan, this chapter outlines the CAKN strategy for generating and handling metadata which serve as the principal documentation for data.

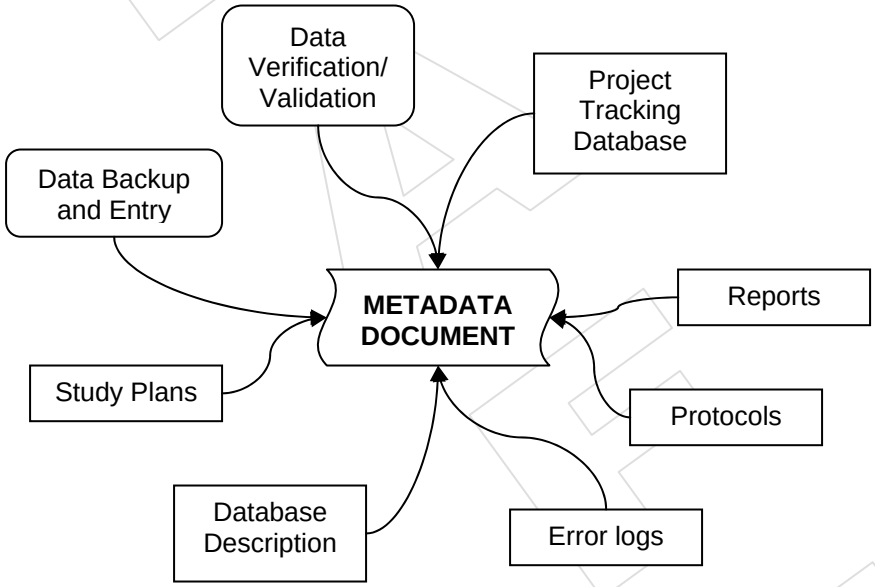


Figure 7.1. Project and data documentation. The metadata document contains all information pertaining to project data.

VII.1. Mandate for Documentation

Executive Order 12906, mandates that federal agencies "...document all new geospatial data it collects or produces, either directly or indirectly..." using the Federal Geographic Data Committee (FGDC) Content Standard for Digital Geospatial Metadata (CSDGM,

www.fgdc.gov/metadata/contstan.html). In addition, EO 12906 directs agencies to plan for legacy data documentation and provide metadata and data to the public.

The FGDC Biological Data Profile (*www.fgdc.gov/standards/status/sub5_2.html*) contains all the elements of the CSDGM and includes additional elements for describing biological data sets. Metadata created in compliance with the Biological Data Profile can be added to the National Biological Information Infrastructure (NBII, *www.nbii.gov/datainfo/metadata*) Clearinghouse. Although not a requirement, completion of the Biological Data Profile for appropriate data sets is recommended.

All GIS data layers must be documented with applicable FGDC and NPS metadata standards. The NPS GIS Committee requires all GIS data layers be described with FGDC standards and the NPS Metadata Profile (*nrdata.nps.gov/profiles/NPS_Profile.xml*). While there are numerous tools available for developing metadata, the NPS Integrated Metadata System Plan (*science.nature.nps.gov/im/datamgmt/metaplan.htm*) recommends three desktop applications: Dataset Catalog, ArcCatalog, and Spatial Metadata Management System (SMMS). Figure 7.2 illustrates the general NPS strategy for creating and distributing metadata.

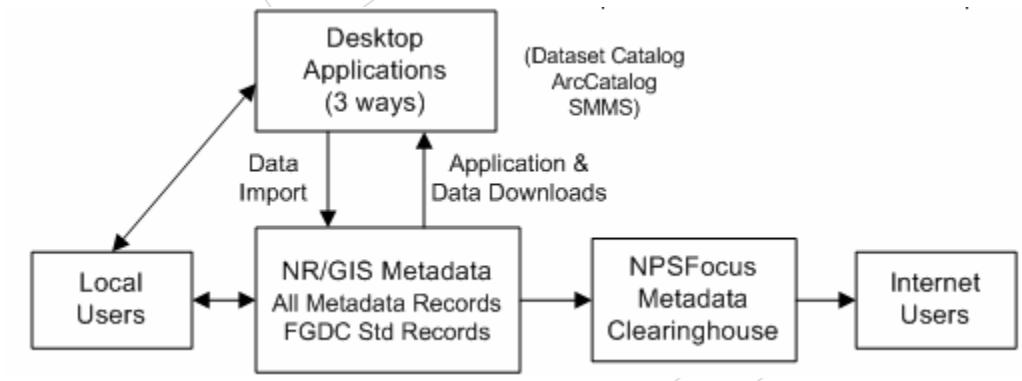


Figure 7.2. NPS Integrated Metadata System. The NR-GIS (NR/GIS) holds metadata records which are publicly accessible via the NPS Focus clearinghouse. Chapter IX presents more information about NPS Focus.

VII.2. Documentation Roles and Responsibilities

Chapter II address general project roles and responsibilities regarding documentation. Documentation duties may be summarized as:

Project Leader	Data Manager
Document project planning, execution and progress in reports, correspondence, project-tracking database, etc.	Facilitate metadata generation by providing tools and assistance in their use
Generate formal dataset metadata	Establish metadata dissemination protocols (cataloguing of metadata on accessible servers)
	Format metadata for dissemination

The duties for the project leader represent a broad scope of tasks, including direction of project personnel and general project management; all aimed at producing a comprehensive metadata document.

The Network data manager is cast in the unwritten role of motivator or prodder to those with metadata generation responsibilities. The data manager ensures that metadata are available in formats suitable for distribution (see Metadata Formats below).

VII.3. Documentation Process

Metadata Tools

As noted above, the NPS Integrated Metadata System Plan recommends three desktop applications for generating metadata:

1. Dataset Catalog (developed by the I&M Program)
2. ArcCatalog (commercial metadata tool)
3. SMMS (commercial metadata tool)

Appendix H contains brief descriptions of these applications as well as a fourth tool, the Metadata Parser (mp) used to check metadata completeness and generate various metadata output formats. The CAKN will utilize a combination of SMMS and ArcCatalog (as well as the Meta Parser) to create metadata and related output.

STRATEGY
The CAKN will utilize a combination of SMMS and ArcCatalog (as well as the Meta Parser) to create metadata and related output.

The end goal, however, is a formal, comprehensive metadata document that meets federal (FGDC) and NPS standards. The tools and means of achieving this end product are numerous and changing. Project leaders may use any tool deemed suitable for the project that meets program requirements.

Metadata Process

Figures 2.1 and 4.4 indicate the points within the standard project workflow for metadata development. Metadata development begins with project design and planning (Figure 2.1). The CAKN data manager will establish a standard operating procedure for metadata generation and maintenance.

STRATEGY
The CAKN data manager will establish a standard operating procedure for metadata generation and maintenance.

STRATEGY
In general, a single metadata document will apply to both raw and certified versions of the data. Metadata records will be stored with both hard copy and digital archive data.

In general, a metadata document will be initiated using ArcCatalog, SMMS or Dataset Catalog in the design and planning stage of a project. As the project progresses, this document will be augmented by the project leader to included relevant project details. Full development of metadata for project data will be completed after the dataset is certified by the project leader. Once metadata are complete, the Network data manager saves the document in different formats (Table 7.1) and parses the information into varying levels of information (Table 7.2). Metadata are saved on the

primary CAKN server along with data and copied to the I&M Program NR-GIS metadata server (Figure 7.3). Hypertext links within the metadata document will point to accompanying datasets served on the primary CAKN server. In general, a single metadata document will apply to both raw and certified versions of the data. Section VII.9, below, further specifies that derived data (typically data resulting from the analysis of certified data) may require a separate metadata document. Metadata records will be stored with both hard copy and digital archive data.

VII.4. Metadata Formats

Table 7.1 lists the metadata file formats to be saved on the primary CAKN server:

Table 7.1. Metadata file formats.

Format	Purpose
HTML	Standard format for web browsers
FAQ HTML	Frequently Asked Questions version of metadata
ASCII	Standard text file
XML	Extensible Markup Language for application programming

These file formats will be generated using the Meta Parser (Appendix H) and will be viewable on the CAKN website. The HTML file version will server as the primary file for display on the website and internal hyperlinks will allow for easy viewing, saving and printing of the other formats.

VII.5. Metadata Parsing

Metadata for primary server CAKN data will conform to FGDC guidelines and be parsed into three nesting levels of detail each designed with a specific audience in mind (Table 7.2).

Table 7.2. Metadata parsing strategy.

Metadata Parsing	Purpose
Level 1, or “Manager Level”	Overview of the dataset crafted to impart quickly the essentials needed to understand the product.
Level 2, or “Scientist Level”	Additional details that allow for rapid scientific assessment of the product.
Level 3, or “Full Metadata”	All components of supporting information such that the data may be confidently manipulated, analyzed and synthesized.

Parsed metadata will be viewable and printable via the CAKN website. The XML version of the metadata document will serve as the basis for parsing metadata.

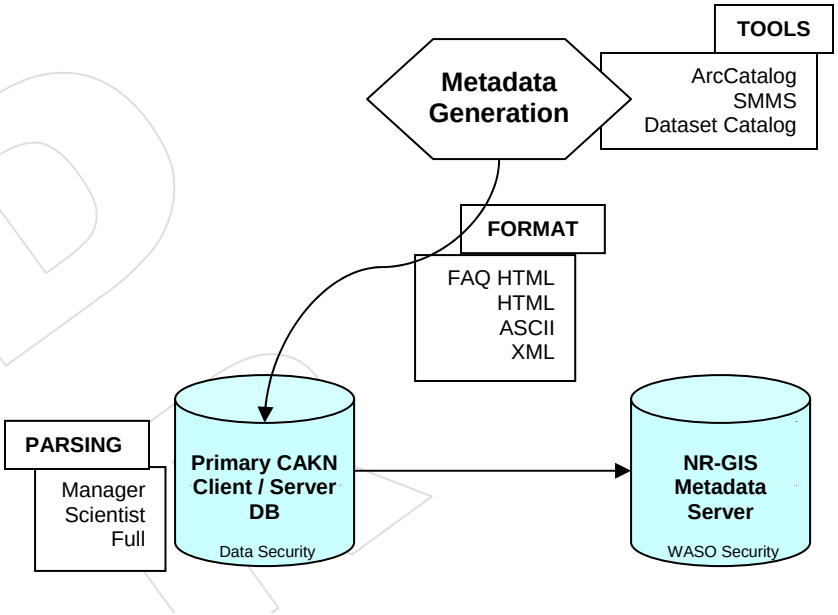


Figure 7.3. Metadata repositories for the CAKN. The CAKN server acts as primary storage of metadata. Metadata are available publicly via the NR-GIS Metadata Server.

VII.6. Metadata Maintenance

The CAKN SOP for metadata generation and maintenance (see Section VII.3) will provide details on the mechanics of handling metadata. Generally, metadata will be created by the project leader, with assistance from the data manager, via an application interface tied to project data on the primary CAKN server. The Network data manager ensures that updated metadata are copied to the I&M Program NR-GIS server. The NR-GIS server will hold XML metadata which will be parsed for viewing in applications on that server.

VII.7. Protocol Versions

Chapter VI noted that database design for CAKN data will include a means for recording the protocol version under which each piece of data was collected (see Section VI.5). The metadata document will contain information about protocol versions used to collect the data.

VII.8. Non-Program Data Documentation

STRATEGY
Any contracts entered into by the CAKN with data producers will stipulate the submission of FGDC-compliant metadata in HTML, ASCII or XML format.

Data which are generated and/or managed outside the I&M Program but used in analysis with CAKN data (see Chapter V) or distributed in any manner by the NPS require the same level of documentation produced for CAKN-generated data. This includes data produced under contract with the NPS. Metadata for non-program data will be requested from the originating entity by the project leader. Any contracts entered into by the CAKN with data producers will stipulate the submission of FGDC-compliant metadata in HTML, ASCII or

XML format or in a format specified by the project leader and approved the Network data manager. The Network data manager will assist the project leader in metadata acquisition by providing tools, format protocols and file transfer services.

VII.9. Derived Data Documentation

Several of the CAKN vital sign monitoring projects will produce datasets derived from originally collected and certified datasets. When a derived dataset becomes a data product, metadata are required for that dataset. Like regular project data, derived data products may not be released without proper documentation (see Chapter IX, Section IX.7). Metadata for any such derived data will include a data lineage pointing to parent datasets.

When a derived dataset becomes a data product, metadata are required for that dataset.

VII.10. Legacy Data Documentation

Data acquired via data mining or other means does not enjoy a level of stewardship afforded project data. When older data that have no current steward (typically a project leader) are found to be useful, metadata for those data must be established. Also, as noted in Chapter IX, Data Dissemination, no data may be released without accompanying metadata. If ancillary data are used in the creation of derived data products (used in analytical comparison to project data), metadata for those data are necessary to fully document the derived data product.

If ancillary data are used in the creation of derived data products, metadata for those data are necessary to fully document the derived data product.

VIII. Data Analysis and Reporting

In this chapter, we describe approaches to how data collected by the CAKN will be analyzed, including who is responsible and how often analysis will occur. We also describe the various reports and other products of the monitoring effort, including what they will include, who the intended audience is, how often they will be produced and in what format, and who is responsible for these products. Because we are actively planning for monitoring data acquisition, this chapter builds analysis and reporting approaches around anticipated monitoring data. However, general concepts described in this chapter may be applied to any CAKN project.

In summary, the CAKN strategy towards data analysis and reporting rests upon providing sufficient funding for these activities so that they occur promptly—that is, to report on the previous phenological year (Oct-Sept) by the following March. The CAKN will also focus on producing an annual integrated "State of the Parks" report that effectively communicates the changes and trends observed in each Vital Sign to our primary audience—the natural resource managers of each park.

Most of the material here is taken from the CAKN Monitoring Plan (MacCluskie and Oakley, 2004).

VIII.1. Data Analysis

For the purposes of this program, we define *data analysis* as the processes by which observations of the environment are turned into meaningful information. We have defined *data analysis* broadly to include all evaluations of data after the data are collected and entered into an electronic file. Thus, data analysis includes quality control checks that occur during summarization and exploratory data analysis and extends through to analytical procedures leading to conclusions and interpretations of the data. We present some general considerations on analysis of monitoring data and outline the general strategy that CAKN will take for all Vital Signs. We also describe the specific approaches currently planned for each Vital Sign.

Analysis of Monitoring Data—General Considerations and CAKN Strategies

Monitoring data pose challenges to analysis because of inherent temporal associations in the estimates. It is essential that we use statistical analyses that accommodate these associations. These approaches include time series analyses, longitudinal data analysis (including repeated measures), trend estimation (many methods), direct estimation of change, and cumulative summary (CUSUM) techniques. Application of these analytical methods will require working closely with statisticians throughout the initial design process and during subsequent analyses of program data.

Application of these analytical methods will require working closely with statisticians throughout the initial design process and during subsequent analyses of program data.

Many of the difficulties typically encountered in analysis of monitoring data can be avoided by proper planning, including the use of probability sampling designs. Appropriate analysis of monitoring data is directly linked to the monitoring objectives, the spatial and temporal aspects of the sampling design used, and management uses of the data. Analysis methods need to be considered when the objectives are identified and the

sampling design is selected, rather than after the data are collected. Failure to adequately consider analysis methods during monitoring program development could result in use of sampling designs that are either inadequate or too complex to meet the monitoring objectives. The purpose of this chapter, and of the specific Standard Operating Procedures (SOPs) on data analysis for each CAKN Vital Sign, is to ensure that the sampling designs and analysis methods we plan to use will allow us to meet our monitoring objectives.

The purpose of this chapter, and of the specific Standard Operating Procedures (SOPs) on data analysis for each CAKN Vital Sign, is to ensure that the sampling designs and analysis methods we plan to use will allow us to meet our monitoring objectives.

The network has developed several strategies to guide the development of SOPs for the data analysis for the CAKN Vital Signs program. These strategies include the use of straightforward (equal probability) sampling designs for as many Vital Signs as possible (see CAKN Monitoring Plan, Chapter 4). Sampling designs that are highly structured (i.e., include many stratifications) make subsequent analyses difficult. Having unstructured designs is important in long-term monitoring because it allows more flexibility in the analysis phase (Overton and Stehman 1995, 1996, Nusser et al. 1998). Another CAKN strategy is to work closely with statisticians in developing and implementing change detection analyses. The network is establishing a multiyear agreement with biometricians at the University of Alaska Fairbanks, Institute of Arctic Biology for this purpose.

STRATEGY
Data analysis SOP development will:

- Use straightforward, minimally structured sampling designs.
- Work closely with statisticians in developing and implementing change detection analyses.
- Provide adequate support to project leaders for data analysis (including hiring of staff).

A central tenet of the CAKN program is that data will be analyzed and reported promptly. Parks need to be alerted to changes in park ecosystems as soon as the changes can be detected—not several years after the fact. Thus, it is imperative that monitoring data be analyzed and reported on as soon as possible. Mechanisms to support prompt analysis and reporting have been built into the data management plan (i.e. data must be entered into the database within 1 month of returning from the field). Additional mechanisms will be established in the Data Analysis SOP for each Vital Sign to support prompt analysis and reporting.

A central tenet of the CAKN program is that data will be analyzed and reported promptly.

One of the primary problems leading to long delays in analysis and reporting is a lack of explicit funding for this activity (Caughlan and Oakley 2001). Thus, the CAKN strategy includes providing adequate support to project leaders for data analysis. This will typically include hiring of staff to provide support so the project lead has the time needed for analysis.

The first step in analysis is summarization (Mulder et al. 1999). This step is a critical part of overall quality control. The data need to be summarized promptly to identify missing values, outliers and other problems related to data collection procedures and the data entry process (Jeffers 1994, Reid 2001). Routines for summarization will be prepared and codified in the SOPs for each Vital Sign. The exact form of the summaries will vary depending on the Vital Sign. In general, however, the approach will include use of graphical techniques to show the data in space and time, using measures of central tendency and variation.

The second step in analysis of CAKN data sets will employ an analysis method that allows us to immediately determine if something has occurred that is out of the bounds of expected variation. Under consideration for this use are the conformance metric developed by Debevec and Rexstad (2000) and the CUSUM approach (Manly and Mackenzie 2003). The conformance metric separates out sampling variation from total variation to provide a measure of the natural variation in an attribute due to ecological processes. Once we establish a baseline to characterize "normal" variability, we can view new observations of the attribute and determine how well they conform to the documented history of the attribute. The conformance metric is the probability that a new observation comes from the same underlying process as the baseline. Hence, a small conformance indicates a change. Using conformance as a metric of change allows information from each Vital Sign to be translated to a common reporting system (i.e., *is everything going about as expected or not?*) and can be pooled hierarchically to any desired level. In the similar CUSUM approach, charts are created that allow systematic deviations to be easily seen. Both approaches are relatively easy to carry out and can complement other approaches to analysis of changes and trend. See Chapter VI, Quality Assurance and Control for precautions regarding the detection of outlier data.

The third step in analysis of CAKN data sets will be in-depth analyses of change over time. Specific methods of change, trend, or temporal pattern detection for each Vital Sign will be used and reported at pre-determined intervals. When appropriate, other analyses such as species-habitat relationships or community ordinations may also occur.

Analysis steps:

1. *Summarization*
2. *Outlier detection*
3. *Change over time*
4. *Vital Sign relationships*
5. *Time series analysis*

The main approaches we currently intend to use for trend detection are time series analysis (Brockwell and Davis 2002), for climate attributes, and mixed linear models (Diggle et al. 1994, Pinheiro and Bates 2000) for most other attributes. Mixed linear models use information from the variance and covariance structure of the data to reduce correlations typical in repeated measures and time series data.

We expect the analysis methods used in the program to change over time. During the first 5 to10 years of the program, the focus will be on summary of findings for a given year across the spatial scale of the network. Comparisons to previous years will be

made if data are available. Once measurements have been made over three points in time, conformance can be calculated and analyses of trend can begin. After measurements have been made for longer periods, modeling of relationships among Vital Signs can begin, and time series analyses can be approached.

A variety of analytical software programs are available for data analysis. The CAKN is using R, a programming language and environment for statistical computing and graphics, as a primary arena for data analysis (Maindonald and Braun 2003). Customized R functions can be written to perform data analysis, generate graphical displays, and automate repetitive reports. Packaged R routines can be run locally on a workstation or delivered over the web (Debevec and Rexstad 2004). R is open-source software available at no cost. The network has started working in R for summarization of data for the "Vegetation Structure and Composition" and "Passerine Birds" Vital Signs, and will continue to work in this direction.

Initial Analysis Approaches for CAKN Vital Signs

The initial analysis approaches to be used for each Vital Sign to be monitored in the first phase of CAKN program implementation are shown in Table 8.1. We also identify, for each Vital Sign, the person who has the lead responsibility for data analysis. In some cases, the analysis may be conducted by a person outside of the National Park Service. In all cases, the person within the NPS designated to conduct the analysis or manage the agreement under which another person conducts the analysis, is identified. In writing the standard operating procedures for data analysis for each CAKN Vital Sign, we have attempted to provide as much detail as possible about the initial steps of data analysis. At some point, however, the steps in analysis cannot be prescribed *a priori*, and we have therefore described suggested approaches that would be appropriate given the objectives and sampling designs used.

For analysis of climate data, we are currently working with the Western Region Climate Center to develop routine analyses for regular reporting. Analysis of snowpack data is conducted under the auspices of the Natural Resources Conservation Service according to established procedures, and the data are posted at the Western Region Climate Center. All air quality data from Denali are analyzed under the auspices of the NPS Air Quality Program according to their established procedures.

Current estimates of vegetation parameters that apply to the sampled areas in each park will be based on a 6-year moving average of individual year estimates. Current estimates that apply to the sampled areas can be produced after one field season. Trends in vegetation parameters will be estimated using mixed linear models that potentially contain effects for year, site, and external covariates such as elevation. Spatial and temporal correlation in responses will be considered in the mixed linear model, and results will be adjusted if necessary. Low precision trend estimates can be produced after two field seasons. High-precision trend analyses can be produced beginning in year 7 after the first rotation of sampling effort to all mini-grids is complete.

A similar approach will be used for analysis of the passerine bird data, also collected within the minigrid design. The revisit design for the passerine bird data has more structure in it to account for higher interannual variability. Another important feature of the passerine bird data analysis is the use of distance estimation to account for differences in species detectability.

Analysis methods for golden eagles, peregrine falcons, moose and wolves will follow standard procedures in practice now for these long-standing studies. For Golden Eagles and Peregrine Falcons, analysis methods are straightforward for tracking nesting territory occupancy and productivity. A new aspect to the Golden Eagle protocol is being added to assess sightability. Depending on results, the protocol could be revised to produce an estimate, rather than a count. For moose, an estimated population size is calculated using established programming (SMOOSE) developed by Ver Hoef (2001) and available on the internet from the Alaska Department of Fish and Game (*winfonet.alaska.gov*). The estimates are available immediately after the survey. For wolves, radio-tracking occurs throughout the year, and a home range program (Hooge and Eichenlaub 2000) is used to map pack territories. Population and density estimates are produced twice a year—in October at the beginning of winter, and in April, at the end of winter.

Table 8.1. Summary of data analysis approaches and responsibilities for each Vital Sign included in the initial Central Alaska Network Vital Signs monitoring program.

Vital Sign			Data Analysis Approach	How Often?	Who is Responsible for Data Analysis?
Physical Drivers	Climate		Time Series Analysis	Data available in real-time; summary analyses performed annually	NOAA-National Weather Service, Western Region Climate Center under terms of agreement managed by Denali Environmental Specialist, Pam Sousanes
	Snowpack		Mixed Linear Models	Annual	USDA-Natural Resource Conservation Service, Rick McClure, under terms of agreement managed by Denali Environmental Specialist, Pam Sousanes
	Water Quality (Ponds)		Mixed Linear Models	Annual	Yukon Charley Rivers-Gates of the Arctic Aquatic Ecologist, Amy Larsen
	Air Quality		Time Series Analysis	Annual	NPS National Air Quality Monitoring Program
Veg.	Structure and Composition		Mixed Linear Models	Annual	Denali Botanist, Carl Roland
Fauna	Birds	Passerines	Distance estimation to address detection bias; Mixed Linear Models	Annual	Denali Wildlife Biologist, Carol McIntyre
		Golden Eagles (DENA)	Mixed Linear Models	Annual	Denali Wildlife Biologist, Carol McIntyre
		Peregrine Falcons (YUCH)	Mixed Linear Models	Annual	Yukon-Charley Rivers Wildlife Biologist, Nikki Guldager, working initially with Skip Ambrose.
	Large Mammals	Moose	Mixed Linear Models	Every 3 years (for each park)	Yukon-Charley Rivers Wildlife Biologist, John Burch
		Wolves	Use capture-recapture data to generate 95% minimum convex territories for each park		Denali Wildlife Biologist, Tom Meier

VIII.2. Reporting

Communicating the findings of the monitoring program is reporting. In this section of the Data Management Plan, we begin by discussing general considerations about reporting and identifying general CAKN strategies about reporting. We then identify the main methods we will use for reporting to specific audiences, as well as the specific reports to be generated.

Reporting Monitoring Data—General Considerations and CAKN Strategies

Reporting is critical to the long-term success of the CAKN Vital Signs program. Results must be credible and delivered in a timely fashion to the appropriate audiences in a manner that is understandable to them. There are multiple audiences for monitoring data produced by the CAKN Vital Signs program, and each requires information formatted and presented in specific ways. The main audience for monitoring data is the resource managers of each network park and other managers in the National Park Service system, who will use the information to assist with their management decisions.

The main audience for monitoring data is the resource managers of each network park and other managers in the National Park Service system, who will use the information to assist with their management decisions.

Although making monitoring findings available to resource managers and other audiences is the underlying reason for monitoring programs, failure to report or long delays in reporting are common problems. Sometimes the reasons for not reporting do not lie in reporting mechanisms per se, but are the result of problems earlier in the monitoring process (i.e., setting measurable objectives, sampling design, feasibility of carrying out the work, data management, data analysis). Thus, for the reporting end of the monitoring program to work well, all other parts of the monitoring program must also be functioning properly. As with data analysis, reporting is an activity that needs to be adequately funded so that reports are produced on schedule (Caughlan and Oakley 2001). Too often, reports are delayed while the next cycle of data collection takes place.

For the reporting end of the monitoring program to work well, all other parts of the monitoring program must also be functioning properly.

Producing reports that effectively communicate findings from the monitoring program is also critical. Oakley (2004) reviewed monitoring reports from a variety of programs and observed that key results often are buried in text and would be difficult for a busy manager to find. Monitoring reports can also be quite lengthy and difficult to read in their entirety. Thus, an important component of producing effective monitoring reports is to improve presentation of results. In this regard, increasing use of visual methods for communicating results (i.e., graphical techniques) is a key strategy. The network will work towards improving data presentations using some of the graphic techniques suggested by Tufte (1983, 1990, 1997), Cleveland (1993, 1994) and others.

The CAKN vision for reporting includes the following central themes: (1) We will prepare monitoring reports that are understandable and useful to our primary audience: park resource managers, (2) We will prepare reports promptly, and (3) All reports will be readily available. To achieve this vision, the network has adopted the following strategies:

1. The budgets for each Vital Sign will include adequate funding to support the production of required annual and periodic reports.

2. All monitoring data and all reports and information generated from the monitoring data will be made available promptly via the internet, subject to applicable law (see Chapter IX for dissemination information).

3. All written reports will follow the current format guidance set by the Alaska Region Inventory and Monitoring Program (National Park Service, Alaska Region, undated; *www.nature.nps.gov/im/units/AKRO/Documents/Admin/Guidelines/ReportSpec_AK.pdf*).

4. All written reports will include a brief summary that includes the main findings presented in the report, using language understandable to a general audience not conversant with the specific technical details of the subject matter.

5. The use of graphical methods for presenting data following principles described by Tufte (1983, 1990, 1997) and Cleveland (1993, 1994) will be encouraged.

STRATEGY
The CAKN vision for reporting includes the following central themes:

- Monitoring reports will be understandable and useful to our primary audience: park resource managers
- Reports well be prepared promptly
- All reports will be readily available

Initial Reporting Approaches

The list of reports to be produced by the CAKN is based on national guidance, modified to reflect CAKN reporting goals (Table 8.2). For administrative reporting, the network will rely on the "Annual Administrative Report and Workplan" required to be prepared in the fall of each year. For reporting of monitoring results, the network will use a variety of annual and periodic written reports, a biennial conference for the network, and participation in other scientific forums (i.e., scientific meetings, symposia, etc.). The network will also conduct periodic program and protocol reviews.

As discussed in the CAKN Monitoring Plan, the CAKN has structured its Vital Signs Monitoring program around a holistic ecosystem model and has focused on creating an integrated program. The vision of an integrated program will be carried through in the reporting stage by the annual production of a "State of the Parks" type report for the network. Initially, this report will be constructed from the summaries provided in the annual reports produced for each Vital Sign. The report will be short and will emphasize graphical summaries of the data. We will work toward incorporating conformance measures for each Vital Sign as an initial method of integrating monitoring findings. The first "State of the Parks" report will be produced in March 2006, following the first full year of program implementation.

As a network of subarctic parks, the CAKN annual work schedule is strongly tied to the annual climate cycle. The CAKN parks are typically covered by snow for ~8 months of the year and snow-free for only ~4 months. Although a majority of the field work for most Vital Signs occurs during the snow-free months, some Vital Signs are measured year-round, and others occur mainly in fall and winter. The differences in timing of the main field work leads to some challenges in scheduling of annual reporting and in producing a report that integrates all Vital Signs. Because the program is ecological, we have decided to use the phenological year (starting from freeze-up in October to following September) as the basis for reporting. All annual reports will be produced in March to describe conditions and changes occurring in the previous phenological year. The annual reports may also include available data for the current phenological year (i.e., fall caribou counts, snowfall, breakup prognosis), but the primary focus will be the previous year.

Table 8.2. Reports to be produced by the Central Alaska Network Vital Signs Monitoring Program.

Type of Report	Purpose of Report	Primary Audience	Frequency	Peer Review Process
Annual Administrative Report and Work Plan	Account for funds and FTEs expended; Describe objectives, tasks, accomplishments, products of the monitoring effort; Improves communication within park, network, region, Program;	Superintendents, network staff, regional coordinators, and Service-wide program managers; admin. report used for annual Report to Congress.	Annual; due to WASO by November 8	Review and approval by Regional Office and Servicewide Program manager
Annual Reports for each Protocol or Project	Archive annual data and document monitoring activities for the year; Describe current condition of the resource and provide alert if data are outside bounds of known variation; Document changes in monitoring protocols; Communication within the park or network;	Park resource managers; network staff; external scientists	Annual; published each March	Peer reviewed at network level
Annual Report on "State of the Parks" for the CAKN Vital Signs Program	Describes current conditions of park resources; Report interesting trends and highlights of monitoring activities; Identifies situations of concern; Explores future issues and directions;	Superintendents; Park resource managers; network staff; external scientists; public	Annual; published each March	Peer reviewed at network level
Analysis and Synthesis reports – trends	Determine patterns/trends in condition of resources being monitored; Discover new characteristics of resources and correlations among resources being monitored; Analyze data to determine amount of change that can be detected by this type and level of sampling; Context – interpret data for the park within a multi-park, regional or national context; Recommend changes to management of resources (feedback for adaptive management);	Superintendents, park resource managers, network staff, external scientists	3-5 year intervals for resources sampled annually	Peer reviewed at network level
Program and Protocol Review reports	Periodic formal reviews of operations and results (5 year intervals);	Superintendents, park resource managers, network	5 year intervals;	Peer reviewed at regional or national

Type of Report	Purpose of Report	Primary Audience	Frequency	Peer Review Process
	Review protocol design and products to determine if changes needed; Part of quality assurance – peer review process;	staff, Service-wide Program managers, external scientists	data quality reports more often	level
Scientific journal articles and book chapters	Document and communicate advances in knowledge; Part of quality assurance – peer review process;	External scientists, park resource managers, network staff	Varies	Peer reviewed by journal or book editor
CAKN Vital Signs Monitoring Conference	Review and summarize information on CAKN Vital Signs; Helps identify emerging issues and generate new ideas;	Park resource managers, network staff, external scientists	Biennial; in the spring (Feb, March or April)	Peer reviewed at network level
Other symposia, conferences, and workshops	Review and summarize information on a specific topic or subject area; Communication of latest findings with peers; Helps identify emerging issues and generate new ideas;	Park resource managers, network staff, external scientists	Varies	May be peer reviewed by editor if written papers are published
CAKN contributions to the national "State of the Parks" Report	Describes current conditions of park resources; Report interesting trends and highlights of monitoring activities; Identifies situations of concern; Explores future issues and directions;	Congress, budget office, NPS Leadership, superintendents, general public	Annual	Peer reviewed at national level

IX. Data Dissemination

Data management within the CAKN I&M Program aims to ensure that

- Data are easily discoverable and obtainable
- No data that have not been subjected to full quality control are released
- Distributed data are accompanied by complete metadata which clearly establishes the data as a product of the NPS I&M Program
- Sensitive data are identified and protected from unauthorized access and inappropriate use
- A complete record of data distribution/dissemination is maintained

IX.1. Data Ownership

National Park Service Policy on Data Ownership

The National Park Service defines conditions for the ownership and sharing of collections, data, and results based on research funded by the United States government. All cooperative and interagency agreements, as well as contracts, should include clear provisions for data ownership and sharing as defined by the National Park Service:

- All data and materials collected or generated using National Park Service personnel and funds become the property of the National Park Service.

- Any important findings from research and educational activities should be promptly submitted for publication. Authorship must accurately reflect the contributions of those involved.

- Investigators must share collections, data, results, and supporting materials with other researchers whenever possible. In exceptional cases, where collections or data are sensitive or fragile, access may be limited.

The Office of Management and Budget (OMB) ensures that grants and cooperative agreements are managed properly. Federal funding must be disbursed in accordance with applicable laws and regulations. OMB circulars establish some degree of standardization government-wide to achieve consistency and uniformity in the development and administration of grants and cooperative agreements. Specifically, OMB Circular A-110 establishes property standards within cooperative agreements with higher institutions and non-profit organizations. Section 36 of Circular A-110, “Intangible Property” describes the following administrative requirements pertinent to data and ownership:

> *(a) The recipient* (higher institution or non-profit organization receiving federal monies for natural resource inventory and/or monitoring) *may copyright any work that is subject to copyright and was developed, or for which ownership was purchased, under an award. The Federal awarding agency(ies*) (in this case the National Park Service) *reserve a royalty-free, nonexclusive and irrevocable right to reproduce, publish, or otherwise use the work for Federal purposes, and to authorize others to do so.*

Section 36 also states:

(c) The Federal Government has the right to:

(1) obtain, reproduce, publish or otherwise use the data first produced under an award

(2) authorize others to receive, reproduce, publish, or otherwise use such data for Federal purposes

(d) (1) In addition, in response to a Freedom of Information Act (FOIA) request for research data relating to published research findings produced under an award that were used by the Federal Government in developing an agency action that has the force and effect of law, the Federal awarding agency shall request, and the recipient shall provide, within a reasonable time, the research data so that they can be made available to the public through the procedures established under the FOIA (5 U.S.C. 552(a)(4)(A)).

(2) The following definitions apply for purposes of paragraph (d) of this section:
(i) Research data is defined as the recorded factual material commonly accepted in the scientific community as necessary to validate research findings, but not any of the following: preliminary analyses, drafts of scientific papers, plans for future research, peer reviews, or communications with colleagues. This "recorded" material excludes physical objects (e.g., laboratory samples)…
(ii) Published is defined as either when:
(A) Research findings are published in a peer-reviewed scientific or technical journal; or
(B) A Federal agency publicly and officially cites the research findings in support of an agency action that has the force and effect of law.
(iii) Used by the Federal Government in developing an agency action that has the force and effect of law is defined as when an agency publicly and officially cites the research findings in support of an agency action that has the force and effect of law.

Establishing Data Ownership: Cooperative or Interagency Agreements

To ensure that proper ownership, format, and development of network products is maintained, all cooperative or interagency work must be conducted as part of a signed collaborative agreement. Every cooperative or interagency agreement or contract involving the Central Alaska Network must include OMB Circular A-110 cited under the *Reports and Deliverables* Section of all agreements and contracts. The following shows appropriate language to use when citing Circular A-110:

> "As the performing organization of this agreement, *institution or organization name* shall follow the procedures and policies set forth in OMB Circular A-110."

Cooperative or interagency agreements or contracts must include a clearly defined list of deliverables and products. Details on formatting and media types that

Cooperative or interagency agreements or contracts must include a clearly defined list of deliverables and products.

will be required for final submission must be included. Typical products include, but are not limited to, field notebooks, photographs (hardcopy and digital), specimens, raw data, and reports.

Chapter VIII specifies that all reports generated by or for the CAKN must follow current format guidance set by the Alaska Region I&M Program. Other products resulting from cooperative work must follow Alaska Region I&M formatting specifications (*www.nature.nps.gov/im/units/AKRO/products/Products_regional.htm*).

The following statement must be included in the Reports and Deliverables section of all CAKN agreements and contracts:

> "All reports and deliverables must follow the most recent versions of the Alaska Region product specifications."

Cooperative agreements and contracts should also provide a schedule of deliverables that includes sufficient time for NPS review of draft deliverables before scheduled final submissions.

IX.2. Data Classification: protected vs. public

All data and associated information from I&M activities must be assessed to determine their sensitivity. This includes, but is not limited to, reports, metadata, raw and manipulated spatial and non-spatial data, maps, etc. Network staff must carefully identify and manage any information that is considered sensitive. The Network must clearly identify and define those data needing access restrictions and those to make public.

The Freedom of Information Act, 5 U.S.C. § 552, referred to as FOIA, stipulates that the United States Government, including the National Park Service, must provide access to data and information of interest to the public. FOIA, as amended in 1996 to provide guidance for electronic information distribution, applies to records that are owned or controlled by a federal agency, regardless of whether or not the federal government created the records. FOIA is intended to establish a right for any person to access federal agency records that are not protected from disclosure by exemptions. Under the terms of FOIA, agencies must make non-protected records available for inspection and copying in public reading rooms and/or the Internet. Other records however, are provided in response to specific requests through a specified process. The Department of the Interior's revised FOIA regulations and the Department's Freedom of Information Act Handbook can be accessed at *www.doi.gov/foia* for further information.

In some cases, public access to data can be restricted. Under one Executive Order, Director's Order #66 (draft), and four resource confidentiality laws, the National Parks Omnibus Management Act (16 U.S.C. 5937), the National Historic Preservation Act (16 U.S.C. 470w-3), the Federal Cave Resources Protection Act (16 U.S.C. 4304) and the Archaeological Resources Protection Act (16 U.S.C. 470hh), the National Park Service is directed to protect information about the nature and location of sensitive park resources. Through these regulations, information that could result in harm to natural resources can be classified as 'protected' or 'sensitive' and withheld from public release (National Parks Omnibus Management Act (NPOMA)).

The following guidance for determining whether information should be protected is suggested in the draft Director's Order #66 (the final guidance may be contained in the Reference Manual 66):

- Has harm, theft, or destruction occurred to a similar resource on federal, state, or private lands?
- Has harm, theft, or destruction occurred to other types of resources of similar commercial value, cultural importance, rarity, or threatened or endangered status on federal, state, or private lands?
- Is information about locations of the park resource in the park specific enough so that the park resource is likely to be found at these locations at predictable times now or in the future?
- Would information about the nature of the park resource that is otherwise not of concern permit determining locations of the resource if the information were available in conjunction with other specific types or classes of information?
- Even where relatively out-dated, is there information that would reveal locations or characteristics of the park resource such that the information could be used to find the park resource as it exists now or is likely to exist in the future?
- Does NPS have the capacity to protect the park resource if the public knows its specific location?

Natural Resource information that is sensitive or protected requires the:

- Identification of potentially sensitive resources
- Compilation of all records relating to those resources
- Determination of what data must not be released to the public
- Management and archival of those records to avoid their unintentional release

STRATEGY
Network staff will classify sensitive data on a case-by-case, project-by-project basis.

Classification of sensitive I&M data will be the responsibility of the CAKN staff, the park superintendents, and investigators working on individual projects. Network staff will classify sensitive data on a case-by-case, project-by-project basis. They will work closely with investigators for each project to ensure that potentially sensitive park resources are identified, and that information about these resources is tracked throughout the project.

The Network staff is also responsible for identifying all potentially sensitive resources to project leaders working on each project. The project leaders, whether network staff or partners, will develop procedures to flag all potentially sensitive resources in any products that come from the project, including documents, maps, databases, and metadata. When submitting any products or results, investigators should specifically identify all records and other references to potentially sensitive resources. Note that partners should not

STRATEGY
The project leaders, whether network staff or partners, will develop procedures to flag all potentially sensitive resources in any products that come from the project, including documents, maps, databases, and metadata.

release any information in a public forum before consulting with network staff to ensure that the information is not classified as sensitive or protected.

For example, information may be withheld regarding the nature and/or specific locations of the following resources recognized as 'sensitive' by the National Park Service. According to NPOMA, if the NPS determines that disclosure of information would be harmful, information may be withheld concerning the nature and specific location of:

- Endangered, threatened, rare or commercially valuable National Park System Resources (species and habitats)
- Mineral or paleontological objects
- Objects of cultural patrimony
- Significant caves

Note that information already in the public domain can, in general, be released to the public domain. For example, the media has reported in detail the return of condors to the Grand Canyon. If an individual requests site-specific information about where the condors have been seen, this information, in general, can be released. However, the locations of specific nest sites cannot be released.

IX.3. Access Restrictions on Sensitive Data

CAKN staff are responsible for managing access to sensitive data handled by the Program. All potentially sensitive park resources will be identified and investigators working on network projects will be informed that:

- All data and associated information must be made available for review by network staff prior to release in any format

- Any information classified as protected should not be released in any format except as approved in advance by the National Park Service

The Network Coordinator, NPS project liaison, or Data Manager identifies all potentially sensitive park resources to the project leader for each project. Reciprocally, the project leader must identify any known references to potentially sensitive park resources.

When preparing information into any CAKN repository (see Section IX.5), CAKN staff ensures that all protected information is properly identified and marked. All references to protected information are removed or obscured in any reports, publications, maps, or other public forum.

All references to protected information are removed or obscured in any reports, publications, maps, or other public forum.

Network staff will remove any sensitive information from public versions of documents or other media. They will isolate sensitive from non-sensitive data and determine the appropriate measures for withholding sensitive data. The main distribution applications and repositories developed by the I&M Program, (see Section IX.4) are maintained on both secure and public servers and all records marked 'sensitive' during uploading will only become available on the secure servers. Procedures for assigning a sensitivity level to specific records when uploading to both the NPSpecies and NatureBib databases are given on the following websites:

- *http://science.nature.nps.gov/im/apps/npspp/index.htm*
- *http://www.nature.nps.gov/nrbib/index.htm*

Thus, access to data on sensitive park resources can be limited to network staff or research partners. However, limits to how these data are subsequently released must also be clearly defined. It is crucial that the person uploading records to the online applications (repositories) is familiar with the procedures for identifying and entering protected information.

> *It is crucial that the person uploading records to the online applications (repositories) is familiar with the procedures for identifying and entering protected information.*

IX.4. Dissemination Mechanisms

Data and information will be made available to two primary audiences: public and NPS employees. Table 9.1 lists the data and product repositories the CAKN will use. Data and related products will also be available by request via contacts on the CAKN website. Appendix C provides information on the national-level repositories (NatureBib, NPSpecies and NR-GIS).

Table 9.1. Repositories for CAKN Program data and information.

Item	Repository
Reports (public) digital	NPS Focus, CAKN data server (Fairbanks)
hard copy	Alaska Resources Library and Information Services (ARLIS), park libraries
bibliography	NatureBib
CAKN-generated digital datasets and data products (public, non-sensitive) • Certified data and data products (including photographs) • Metadata	NR-GIS metadata and data store, CAKN data server, NPSpecies, EPA STORET.
CAKN-generated digital project data and info (NPS staff, sensitive) • Raw, validated and analyzed data • Metadata • Submitted reports • Digital photographs • Digital presentations	CAKN data server, backup server (Anchorage). Selected vital sign data will be housed externally (see Table 5.1): ADF&G, WRCC, US Census, US Fish & Wildlife
Project product materials • Vouchers • Specimen	UAM, CAKN or park office (according to project protocol)
Non-product project items (hard copy)	CAKN office

Currently (12/15/2004), the NR-GIS Metadata and Data Store is under development. Until procedures and further guidance become available for the use of this repository, the CAKN will disseminate all data developed as part of its I&M Program via the Network

website. When both repositories are completely operational, the Network will upload all applicable data and information to each of those sites as needed.

CAKN Primary Server

First noted in Chapter IV of this plan, the CAKN will operate a server database housing electronic data and information managed by the program. This relational database will allow staff to browse, evaluate, export, analyze and integrate vital sign monitoring data and information for research, management and reporting purposes. Development of this system is planned in three stages:

- Stage 1
  Databases for each vital sign will be designed in MS Access and ported to MS SQL Server. Existing data browsing and entry windows that have been developed primarily in MS Access as part of the monitoring program development will be used with MS SQL Server. Spatial data saved as shapefiles on a separate server drive will be related to data tables via link tables in MS SQL Server.

- Stage 2
  The data dissemination system will be developed to include a web-interface for data browsing, query and download: an Internet Mapping Service (IMS) serving shapefile-based data and information. An on-line form for completing data requests and/or submitting feedback will be included as part of the web-interface. For park and network staff, an ArcMap interface linking shapefiles and MS SQL Server data tables will augment the web-interface.

- Stage 3
  Conversion of all data to GeoDatabase format for serving via the ESRI Spatial Database Engine on MS SQL Server. ArcMap and IMS applications will be configured to use the GeoDatabase formats of the program data.

As noted earlier in this plan, a significant portion of the CAKN's monitoring vital signs draw from data sources outside the NPS (Table 5.1). Protocols for acquiring data from external sources will be developed for each project to allow comparison and integration with internally generated data.

Non-monitoring data

The distinction between "monitoring" and "non-monitoring" data may be construed as the difference between "active" and "finalized" data. The brunt of CAKN data management planning focuses on monitoring data (see Chapter I), however, in general the CAKN will make every effort to upgrade non-monitoring datasets, such as legacy and inventory data, such that they may be served via the client/server database (MS SQL Server) on the CAKN primary data server. Data formatted in this manner will be more easily browsable and manipulatable by end-users (see Chapter IV, Section IV.5).

STRATEGY
In general the CAKN will make every effort to upgrade non-monitoring datasets, such as legacy and inventory data, such that they may be served via the client/server database (MS SQL Server) on the CAKN primary data server.

If not formatted for MS SQL Server, data will still be housed on the primary CAKN server and be discoverable and obtainable via CAKN dissemination mechanisms. Supporting metadata, reports and other summaries of this class of data will be available on the CAKN website and where possible, users may download finalized datasets bundled with supporting documentation.

Alaska Resources Library and Information Services (ARLIS)

ARLIS provides universal access to natural and cultural resources information and serves as a central library for local, state and federal agencies in Alaska. The park service in Alaska has historically used ARLIS for a variety of information needs and the CAKN will build on this relationship by ensuring that all appropriate program output are catalogued with ARLIS for broad distribution. ARLIS may be accessed at *www.arlis.org*.

NPS Focus

CAKN data and products residing in NatureBib and NR-GIS will automatically be searchable via NPS Focus; a digital library and research station currently under development by the NPS Information and Telecommunications Center (ITC). NPS Focus has been released as an Intranet version only (*focus.nps.gov*). A public version is projected for the near future.

As NPS Focus develops, additional databases and repositories utilized by the network are expected to be searchable through this portal.

Water Quality Data

Water quality data collected to meet regulatory requirements is managed according to guidelines from the NPS Water Resources Division. This includes using the NPSTORET desktop database application at the parks to help manage data entry, documentation, and transfer (Appendix C). The Network oversees the use of NPSTORET according to the Network's integrated water quality monitoring protocol and ensures the content is transferred at least annually to NPS Water Resource Division for upload to the STORET database (Figure 9.1).

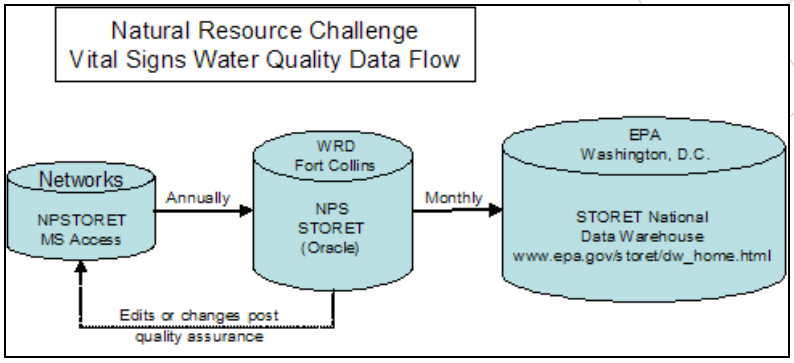


Figure 9.1. Water quality data flow.

IX.5. Data Availability

In general, data will be available upon completion of analysis and reporting. Data for which analysis and reporting has not been completed but are otherwise certified (verified

and validated) will be released no later than one year after certification (see Section IX.6).

STRATEGY
Data for which analysis and reporting has not been completed but are otherwise certified (verified and validated) will be released no later than one year after certification.

Chapter V (Table 5.2) summarizes timing for reporting and integration of CAKN vital sign monitoring data. Data are *integrated* when they are uploaded to dissemination mechanisms outlined in Section IX.4. Chapter VIII presents more details concerning analysis and reporting of CAKN data.

Data release or integration into dissemination mechanisms depends on both the project protocol and project leader. As noted in Chapter VII, no data may be released without proper certification and approval by authorized CAKN staff.

CAKN staff will notify project leaders prior to making datasets available to the public. This will allow each project leader the opportunity to specify in writing any access restrictions that should apply. Any such specifications must indicate how long the data will remain restricted.

STRATEGY
CAKN staff will notify project leaders prior to making datasets available to the public. This will allow each project leader the opportunity to specify in writing any access restrictions that should apply. Any such specifications must indicate how long the data will remain restricted.

IX.6. Data Release Policy

Aspects of the CAKN data release policy are addressed in chapters VI and VII and earlier in this chapter. The CAKN will maintain the following policy regarding the release of data and related products:

1. Only fully documented, certified, non-sensitive data and data products may be uploaded to public distribution repositories or otherwise released to the public or other non-NPS recipient.
2. Any released data or data products must be accompanied by full metadata and any additional supporting documentation as determined by the project leader.
3. CAKN monitoring program data will be made available according to CAKN reporting and data integration schedules specified in the most recent versions of the CAKN Data Management Plan, Chapter VIII and the CAKN Monitoring Plan, Chapter 7 (both titled “Data Analysis and Reporting”).
4. CAKN non-monitoring data will be released upon completion of data certification (verification and validation) and on condition of project leader approval. However,

data for which analysis and reporting has not been completed but are otherwise certified (verified and validated) will be released no later than one year after certification.

5. Distribution instructions for each dataset will be included in respective metadata.
6. Distribution logs specifying recipient name and contact information, intended use of data, export file format, delivery date and method, data content description noting range (by date and geography) of data delivered and description of distributed items (noting that metadata and possibly other supporting documentation were distributed with data), will be kept for each distinct dataset or product managed by the CAKN.
7. At the discretion of the project leader, any reports utilizing the data may also accompany distributed data (note, see Section IX.5 above for additional release information).

STRATEGY
CAKN non-monitoring data will be released upon completion of data certification (verification and validation) and on condition of project leader approval. However, data for which analysis and reporting has not been completed but are otherwise certified (verified and validated) will be released no later than one year after certification.

STRATEGY
Data distribution logs will include:

- Recipient name and contact info
- Intended use of data
- Export file format
- Delivery date and method
- Description of data content
- Description of distributed items (noting that metadata and other supporting documentation were distributed)

IX.7. Feedback Mechanisms

Comments and questions concerning I&M project data are welcome at any time and may be submitted via e-mail or telephone to the primary contact of a project or to the Network Coordinator. The CAKN website will also provide an opportunity for NPS staff, cooperators and the public to provide feedback on data and information distributed as part of CAKN operations. A “comments and questions” link will be provided on the main page of the site for general questions and comments about the Network’s program and projects. A more specific “data error feedback” link will direct comments to CAKN staff pertaining to errors found in website accessible data. Data feedback and response activity will be reported to the Board of Directors and to the Technical Committee on a yearly basis.

Data Error Feedback Response Procedures

The CAKN will use the following procedure to respond to data feedback:

- CAKN staff (or automated process) will immediately acknowledge receipt of any feedback to the sender.
- Data error reports will be recorded in an error log associated with each database (see Section VI.5).
- Data errors will be investigated within one week of submission. If appropriate and feasible, confirmed errors are immediately corrected in all data repositories. If immediate correction is not possible, the data in question are immediately removed from all dissemination mechanisms (repositories, see Table 9.1).
- A report detailing the reported error and response will be prepared by the network data manager and submitted to the project leader, network coordinator and the error reporter. If a data error is not immediately correctable, the error response report will include recommendations for correcting the error.
- An appropriate level of communication to stake holders will be maintained during all data error investigations.

X. Data Maintenance, Storage and Archiving

CAKN data maintenance, storage and archiving procedures aim to ensure that data and related documents and materials (digital and physical) are

- Kept up-to-date with regards to content and format such that the data are easily accessed and their heritage and quality easily learned.
- Physically secure against environmental hazards, catastrophe, and human malice
- Archived in a manner that expedites recovery if needed

X.1. Data Maintenance

All CAKN data maintenance will conform to NPS standards and policy.

Chapter IV presents the basic infrastructure in which CAKN data and information will be maintained. Data maintenance procedures in the CAKN apply primarily to CAKN-generated data. We anticipate most maintenance activity will involve active monitoring datasets however finalized CAKN-project data will also be maintained along with active data in a common relational database system (the CAKN primary data server; see chapters IV and IX). Related metadata as well as database application interface tags will discern between active and finalized data.

We anticipate that a large body of finalized data, such as legacy and inventory data, will not require the same level of maintenance but will nonetheless be either directly accessible or otherwise obtainable via CAKN dissemination mechanisms (see Chapter IX, Section IX.5). Any data, however, formatted for inclusion in the MS SQL Server database, will be generally maintained along with active data in the database.

Digital File Maintenance

Primary digital data maintenance will be performed on the main CAKN server in Fairbanks. The data and information content of CAKN files stored on this server will be kept current. Dataset metadata (see Chapter VII) and error logs (see Chapter IX) will record data maintenance activities. Database maintenance logs will record any changes to project databases (structure, programming, etc.; see Chapter VI, Section VI.5). File naming protocols and system file dates will serve as version control (see Chapter IV).

A catalogue of the data and information on the CAKN server will be maintained on the CAKN website and reflect changes or updates to datasets. These changes or update records will typically reflect data collection, entry, and certification dates but may also record actions to correct errors in the dataset or to migrate data to new digital formats. National and regional repositories for CAKN data and information (see Chapter IX) will be updated to reflect current stores on the CAKN server.

Digital File Types - Maintenance

As noted, actively maintained CAKN data for on-going projects will be stored as MS SQL Server databases (related tables). Specific project protocols may require different or additional file types not compatible with MS SQL Server. These data will be housed on the primary CAKN server within the corporate file structure (see Chapter II).
As noted in Chapter IX, spatial data will initially be maintained as Shapefiles but migrate to the ESRI GeoDatabase model as the CAKN data management system evolves.

Non-monitoring data will be maintained in original formats (project specific) unless revised for inclusion in the MS SQL Server database on the CAKN primary data server.

Digital Backup Plan

The primary CAKN server will be housed in the Fairbanks YUGA office as part of the YUGA local area network (see Chapter IV). All digital files on this server will be backed up daily onto a 1500 gigabyte RAID (Redundant Array of Independent Disks) maintained by YUGA information technology personnel. These data will also be backed up via the regional network to a server in Anchorage (see chapters IV and IX) at least weekly (we anticipate the ability to backup daily in the near future; 12/15/2004). All digital files will be restorable from either the RAID or the Anchorage server in the event of data loss on the primary CAKN server. Both of these backup sources will be read-only and accessible for data restoration purposes only.

STRATEGY
All digital files will be restorable from either the RAID or the Anchorage server in the event of data loss on the primary CAKN server. Both of these backup sources will be read-only and accessible for data restoration purposes only.

Additional, selected data backups will be made to standard media (CD, DVD, HD-DVD, external hard drive) by the network data manager as needed. It is recognized that project leaders will employ (and be encouraged to do) their own data backups as project work progresses.

Hardcopy Data and Information - Maintenance

Hard copy data and information will be maintained according to the archival procedures outlined in Section X.3. Catalogues of archived material will be searchable on the CAKN website and maintained to reflect basic accessioning information for each item. Collections will be maintained according to NPS archiving protocols (*www.nps.gov/policy/DOrders/DOrder24.html*).

X.2. Digital Data Archival

Chapter IV, Data Management Infrastructure, outlines basic data archival within the CAKN (Figure 4.4). Generally, the data maintenance activities described above will serve to archive data maintained on the primary CAKN server in a RAID (Fairbanks) and an offsite backup server (Anchorage). Raw, certified and analyzed data (data products) will be archived (see Chapter IV, Figure 4.4) and a common metadata file will be associated with each. Chapter VII, Data Documentation, addresses metadata requirements for derived data (Chapter VII, Section VII.9).

Finalized Digital Datasets

Datasets that are considered complete and inactive will be saved on the primary CAKN server (and hence backed up, see Section X.1) in both ASCII and native formats. “Native format” is defined as the format (typically “file type”) in which data for a given project are generated. The “native” format may be different for raw, checked and analyzed versions of the data. Note that “specimen” or “sample” data, i.e. physical objects taken from the

field do not fall under this definition of "native" format (see Section X.3). Data are considered complete and inactive when accompanying metadata, as generated by the project leader or other authorized personnel, list the project status as "complete".

Associated digital content such as submitted project reports, photographs, presentations, etc., will be archived along with project data.

STRATEGY
Datasets that are considered complete and inactive will be saved on the primary CAKN server (and hence backed up, see Section X.1) in both ASCII and native formats. Data are considered complete and inactive when accompanying metadata, as generated by the project leader or other authorized personnel, list the project status as "complete".

X.3. Physical Material Archival

Project Products

Chapter IX, Table 9.1 specifies locations for physical items considered project products such as reports. Additional hard copy project product items such as maps, posters, slides, photographs (including aerial), tape recordings, etc., will be stored in the CAKN main office. These items (including project reports) will be archived according to NPS standards (*www.nps.gov/policy/DOrders/DOrder24.html*) and follow the procedures outlined below under "Non-Product Items".

Specimen and Samples

Chapters V (Data Acquisition) and IX (Data Dissemination) address basic field sample and specimen handling procedures. These items will be archived according to NPS standards following procedures provided in project-specific protocols. Unless otherwise specified in the project protocol, specimen and samples will be provided to the network park in which they were collected for curation according to park policy. The CAKN will provide park curators with associated data and material required for curation.

STRATEGY
Unless otherwise specified in the project protocol, specimen and samples will be provided to the network park in which they were collected for curation according to park policy.

Non-Product Items

Non-product hard copy items such as field notebooks and other materials that may be considered raw data or important project information such as correspondence, permits, agreements, etc., will be catalogued and filed in the Network's main office. Non-product items will be archived along with related project material noted above under "Project Products".

Project materials will be consolidated and "packaged" for archival on an annual basis. Project leaders should acquire a project accession number from an NPS archivist. The project leader should package the annual project information as follows:

- A coversheet or memo listing all materials included in the package (see Appendix I for project checklist form).
- All materials are clearly labeled with:
  - Park code
  - Date, or range of dates
- Field notes on acid free paper and stored in a 3-ring binder or book box
- Other paper materials such as reports and data printouts also on acid free paper and placed individually in a acid free, labeled folder
- Specimens, such as plants or fish, are properly labeled
- For specimens not residing in archives, loan paper work is complete and a copy of the form is stored with the project package.
- Read only media (CD, DVD, HD-DVD) of all electronic materials. Disk case should be labeled with:
  - Park Code
  - I&M Network Code
  - Date created
  - Range of dates for information
  - Accession Number
  - List contents

The archive specialist will catalog this information into NPS Rediscovery and arrange for proper storage of the materials.

X.4. Photographs

The CAKN is developing (12/15/2004) a comprehensive photographic database designed to track and manage images of park resources. Photographs from each CAKN project will be entered into this database where attributes such as electronic file name, keywords, project, description, photographer, date and location will be catalogued. Digital photo management will generally follow guidelines established in a draft management strategy prepared by the Alaska Southeast and Southwest I&M Networks (*www.nature.nps.gov/im/units/swan/Libraries/Data_Management/DataManageGuidelines/AKIM_2004_AK_PhotoStrategy_040423.pdf*).

Literature Cited

Caughlan, L. and K.L. Oakley. 2001. Cost considerations for long-term ecological monitoring. Ecological Indicators 1:123-134.

Cleveland, W.S. 1993. Visualizing data. Hobart Press, NJ.

Cleveland, W.S. 1994. The elements of graphing data. Hobart Press, NJ.

Debevec, E. and E. Rexstad. 2000. Measuring the health of the Denali National Park and Preserve ecosystem through combining multiple LTEM datasets. Presentation to Denali National Park and Preserve, Long Term Ecological Monitoring Conference, October 24-25, 2000, Fairbanks, AK.

Debevec, E. and E. Rexstad. 2004. Creating online R functions with deliveR.

Diggle, P.J., K.-Y Liang, and S.L. Zeger. 1994. Analysis of longitudinal data. Clarendon Press.

Hooge, P.N. and B. Eichenlaub. 2000. Animal movement extension to ArcView ver. 2.0. USGS-Alaska Science Center, Anchorage.

Jeffers, J.N.R. 1994. The importance of exploratory data analysis before the use of sophisticated procedures. Biometrics 50:881-883.

MacCluskie, M. and Oakley, K.L. 2004. Vital Signs Monitoring Plan, Phase 3 Report, Central Alaska Network, National Park Service, 121 pp.

Maindonald, J. and J. Braun. 2003. Data analysis and graphics using R – an example-based approach. Cambridge University Press.

Manly, B.F.J. and D.I. Mackenzie. 2003. CUSUM environmental monitoring in time and space. Environmental and Ecological Statistics 10:231-247.

Miller, Abigail B. 2001. Managing data to bridge boundaries. Pp. 316-320 *in* Harmon, David (ed), Crossing Boundaries in Park Management: Proceedings of the 11th Conference on Research and Resource Management in Parks and on Public Lands, Hancock, MI: The George Wright Society.

Mulder, B.S., B.R. Noon, T.A. Spies, M.G. Raphael, C.J. Palmer, A.R. Olsen, G.H. Reeves, and H.H. Welsh. 1999. The strategy and design of the effectiveness monitoring program for the Northwest Forest Plan. Gen. Tech. Rep. PNW-GTR-437. Portland, OR. U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station, 138 pp.

National Park Service, Alaska Region. Undated. Specifications for annual progress reports and final reports submitted to the Alaska Inventory and Monitoring Program, Alaska Region, National Park Service. 7 pp.

Oakley, K.L. 2004. Quest for the "perfect" monitoring report. USGS-Alaska Science Center, Anchorage. Unpublished report, 3 pp.

Overton, W.S. and S.V. Stehman. 1995. Design implications of anticipated data uses for comprehensive environmental monitoring programmes. Environmental and Ecological Statistics 2:287-303.

Overton, W.S. and S.V. Stehman. 1996. Desirable design characteristics for long-term monitoring of ecological variables. Environmental and Ecological Statistics 3:349-361.

Pinheiro, J.C. and D.M. Bates. 2000. Mixed effects models in S and S-PLUS. Springer-Verlag.

Reid, L.M. 2001. The epidemiology of monitoring. Journal of the American Water Resources Association 37(4):815-820.

Tufte, E.R. 1983. The visual display of quantitative information. Graphics Press, CT.

Tufte, E.R. 1990. Envisioning information. Graphics Press, CT.

Tufte, E.R. 1997. Visual explanations. Images and quantities, evidence and narrative. Graphics Press, CT.

Ver Hoef, J.M. 2001. Predicting finite populations from spatially correlated data. Pp. 93-98 in 2000 Proceedings of the section on Statistics and the Environment of the American Statistical Association.

Acknowledgements

The creation of this plan involved many individuals both within and without the National Park Service. The following individuals contributed to this document:

Name	Network	Chapter
John Boetsch	NCCN	II and IV
Rob Daley	GRYN	III
Patrick Flaherty	APHN	IV
Geoff Sanders	NCRN	V
Debbie Angell	SODN	VI
Sara Stevens	NCBN	IX
Margaret Beer	NCPN	X
Karen Oakley	USGS	VIII

Special thanks to Dorothy Mortenson (SWAN) for an early jump start in writing as well as extremely helpful input throughout development of this plan.
Adulations to John Boetsch for coordinating a collective writing effort amongst the network data managers listed above which contributed greatly to the completion of this plan.
The following individuals provided key review and critique: Dr. Maggie MacCluskie (CAKN), Karen Oakley (USGS), Scott Miller (ARCN), Dr. Steve Fancy (WASO), Lisa Nelson (WASO) and Joe Gregson (WASO).
Laura Weaver developed the better-looking figures in this document.

Appendix A: Basic Resource Inventories and repositories for the CAKN.

Inventory	Information Access
Natural resource bibliography	NatureBib (www.nature.nps.gov/nrbib)
Base cartographic data	AK-NPS GIS Team (www.nps.gov/akso/gis)
Geology map	NPS Geologic Resources (www2.nature.nps.gov/geology/inventory/gre_pub2.htm)
Soils map	AK-NPS GIS Team; surficial geol. data (www.nps.gov/akso/gis) and (www.nature.nps.gov/im/units/cakn/DENASoils.htm)
Meteorological data	Western Regional Climate Center (www.wrcc.dri.edu/wraws/akF.html)
Air quality	NPS Air Resources (www2.nature.nps.gov/air/)
Air quality related values	NPS Air Resources (www2.nature.nps.gov/air/)
Water body location and classification	NPS Water Resources (www.nature.nps.gov/water) and AK-NPS GIS Team; hydro data (www.nps.gov/akso/gis)
Water quality data	NPS Water Resources (www.nature.nps.gov/water)
Vegetation map	CAKN (www.nature.nps.gov/im/units/cakn/PlantInventory.htm)
Documented species occurrence of vertebrates and vascular plants	NPSpecies (science1.nature.nps.gov/npspecies) and CAKN (www.nature.nps.gov/im/units/cakn/Inventory.htm)
Species distribution and status of vertebrates and vascular plants	CAKN (www.nature.nps.gov/im/units/cakn/Inventory.htm)

Appendix B: CAKN Data Management Plan Revisions

The latest version of the DMP is available on the CAKN website (*www.nature.nps.gov/im/units/cakn/DataMgt.htm*) and will include a revision log as an appendix.

Revision and maintenance of the DMP will be incremental as needed to reflect best practices and current information. The data management work group (See Section III.2) for the CAKN will be responsible for this level of on-going maintenance. Recommendations for changes may be forwarded to any member of the data management work group by any interested party or user of CAKN Inventory and Monitoring data. These recommendations will be discussed at regular meetings of the data management work group. Changes to the DMP will be made at the discretion of the data management workgroup only after thorough review and assessment of anticipated ramifications. The amount of time required for adequate review and assessment of ramifications will likely vary greatly in accordance with the nature of the suggested change. When enough revisions have been made to merit an additional round of review, this process will be taken up with the network technical committee. Otherwise, the plan will be scheduled for a more formal revision and review on a regular basis (at least every 5 years).

The titles of updated or revised sections of the plan will be followed by brackets containing the date of the latest version of that section and a brief update history. For example, the title of Section II.2 may be followed by "[04/01/2005; Update History: 1/15/2005, 10/23/2004]". This indicates that the current wording of that section was last updated on April 1, 2005 and that it was previously revised on January 15, 2004 and October 23, 2004. The list of previous dates will be limited to two dates followed by a "…" if there are additional dates. A "Document History" will be maintained as part of Appendix A (this appendix) of the Data Management Plan which will provide further details.

Document History:
7/2/04 Doug Wilder
First draft of sections I through III submitted to Maggie MacCluskie, Karen Oakley, Dorothy Mortenson, Scott Miller, and Bill Eichenlaub for review. Initial draft of plan draws largely from work by Dorothy Mortenson and John Boetsch (NCCN). Doug Wilder is taking a lead role in formulating the initial plan.

Appendix C: National-level Inventory and Monitoring Information Management Strategies

The need for effective natural resource information management cuts across NPS divisional boundaries and management strategies must be defined at the highest level possible. In this context, integrated inventory and monitoring of natural resources is multidisciplinary and requires national-level, programmatic data and information management strategies for success.

The basic strategy of natural resource and therefore inventory and monitoring information management is to provide integrated natural resource databases and information systems that enhance NPS managers' and staff's access and use of timely and valid data and information for management decisions, resource protection, and interpretation. Inventory and monitoring information needs are broadly separated into two categories:

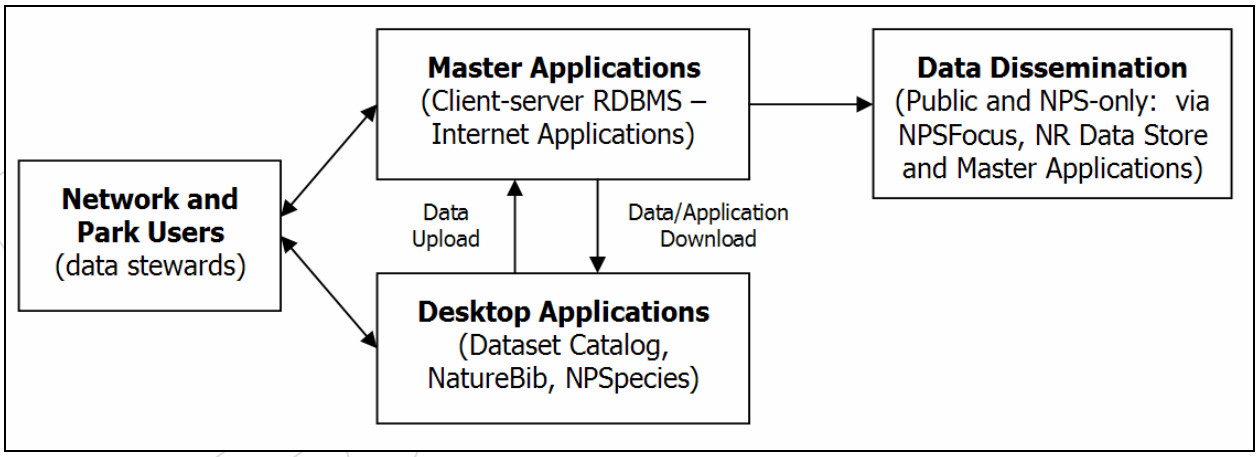
- *Detailed data and information needed for onsite resource management and protection*. The information used to guide natural resource management decisions must be specific to inform and be useful to management staff at parks and central offices.
- *Summary information needed to describe the resources and their condition*. This kind of information usually needs to be aggregated across the National Park Service for use by NPS and DOI managers and central office personnel to answer requests from Congress and for budget, program, and project planning.

The NPS Natural Resource Program Center (NRPC) and the I&M Program actively develop and implement a national-level, program-wide information management framework. NRPC and I&M staff integrate desktop database applications with internet-based databases to serve both local and national-level data and information requirements. NRPC staff members work with regional and support office staff to develop extensible desktop GIS systems that integrate closely with the database systems. Centralized data archiving and distribution capabilities at the NRPC provide for long term data security and storage. NRPC sponsors training courses on data management, I&M techniques, and remote sensing to assist I&M data managers with developing and effectively utilizing natural resource information.

National-level application architecture

To achieve an integrated information management system, three of the national-level data management applications (NatureBib, NPSpecies, and NR-GIS Metadata Database) utilize a distributed application architecture with both desktop and internet-accessible (master) components (Figure 4.2).

Model of the national-level application architecture.



NatureBib

NatureBib is the master database for bibliographic references that merges a number of previously separate databases such as Whitetail Deer Management Bibliography (DeerBib), Geologic Resource Bibliography (GRBib), and others. It also contains citation data from independent databases like NPSpecies and the Dataset Catalog and NR-GIS Metadata Database. It currently focuses on natural resource references, but may eventually be linked to references on cultural resources and other park operations. As with NPSpecies and NR-GIS Metadata Database, it is possible to download data from the master web version into the MS Access desktop version that can be used locally on computers with limited internet connectivity *(http://www.nature.nps.gov/nrbib)*.

NPSpecies

NPSpecies is the master species database for the NPS. The database lists the species that occur in or near each park, and the physical or written evidence for the occurrence of the species (e.g., references, vouchers, and observations). Taxonomy and nomenclature are based on ITIS, the interagency Integrated Taxonomic Information System. The master version of NPSpecies for each park or network can be downloaded from the master website into an MS Access version of NPSpecies. The internet-based version is the master database, which can be accessed via password-protected logins administered by park, network and regional data stewards assigned for each park and network. The master database requires that species lists are certified by networks before any data will be available to the public. NPSpecies is linked to NatureBib for bibliographic references that provide written evidence of a species' occurrence in a park and will be linked to NR-GIS Metadata Database to document biological inventory products. The MS Access application and additional details can be found at the NPSpecies website *(http://science.nature.nps.gov/im/apps/npspp/index.htm).*

Biodiversity Data Store

A digital repository of documents, GIS and other data sets that contribute to the knowledge of biodiversity in National Park units, including presense/absence, distribution and abundance (*http://science.nature.nps.gov/im/inventory/biology/index.htm*).

Dataset Catalog and NR-GIS Metadata Database

Dataset Catalog is a desktop metadata database application developed by the I&M Program to provide a tool that parks, networks, and cooperators can use to inventory and manage data set holdings. Although not designed as a comprehensive metadata tool, the Dataset Catalog is used for cataloging abbreviated metadata about a variety of digital and non-digital natural resource data sets. The Dataset Catalog helps parks and networks begin to meet Executive Order 12906 mandating federal agencies to document all data collected after January 1995. It provides brief metadata and a comprehensive list about all resource data sets for use in data management, project planning, and more stringent metadata activities. As with other service-wide applications, the master metadata database (NR-GIS Metadata Database) is available through a website and will be linked to NPSpecies (the NPS species database) and NatureBib (the bibliographic database). It will be possible to download a version in MS Access format from the master website (*Dataset Catalog: http://science.nature.nps.gov/im/apps/datacat/index.htm and NR-GIS Metadata Database: http://science.nature.nps.gov/nrdata).*

Other National-Level Inventory and Monitoring Information Management and GIS Applications

NPSTORET

STORET is an interagency water quality database developed and supported by the Environmental Protection Agency's (EPA) to house local, state, and federal water quality data collected in support of managing the nation's water resources under the Clean Water Act. STORET is used by NPS as a repository of physical, chemical, biological, and other monitoring data collected in and around national park units by park staff, contractors, and cooperators. The NPS operates its own service-wide copy of STORET and makes periodic uploads to the EPA STORET National Data Warehouse so that data collected by and for parks will be accessible to the public. NPS Director's Order 77 indicates that the NPS should archive water quality data in STORET, and the NPS Water Resources Division (WRD) requires that any data collected as part of a funded WRD project get archived in STORET. NPSTORET (also known as Water Quality Database Templates) the NPS master database designed to facilitate park-level standardized reporting for STORET. The database is still in development, but metadata, protocols, data dictionaries, and reporting capabilities are available through a front-end form. Upon implementation, network staff and cooperators will be able to use the MS Access version of NPSTORET either as a direct database for data entry and management, or as a means of submitting data for upload to STORET by WRD staff. The MS Access application and additional details can be found at: *http://www.nature.nps.gov/water/infodata.htm.* Additional information on STORET can be found at: *http://www.epa.gov/storet.*

Natural Resource Database Template

The Natural Resource Database Template (NRDT) is a flexible, relational database in MS Access for storing inventory and monitoring data (including raw data collected during field studies). This relational database can be used as a standalone database or in conjunction with the GIS software (e.g., ArcView or ArcGIS) to enter, store, retrieve, and otherwise manage natural resource information. The template has a core database structure that can be modified and extended by different parks and networks depending on the components of their inventory and monitoring program and the specific sampling protocols they use. Natural Resource Database Template is a key component of the

I&M program's standardized monitoring protocols. These monitoring protocols include separate modules detailing different aspects of monitoring project implementation, from sampling design to data analysis and reporting, and include data management components that describe database table structure, data entry forms and quality checking routines. Approved monitoring protocols, including the databases that are based on the Database Template, are made available through a web-based protocol clearinghouse (see below). A description of the Database Template application, a data dictionary, and example implementations are located on the NR Database Template website *(http://science.nature.nps.gov/im/apps/template/index.htm).*

Natural Resource Monitoring Protocols Clearinghouse

The Natural Resource Monitoring Protocol Clearinghouse (i.e., Protocol Database) is a web-based clearinghouse of sampling protocols used in national parks to monitor the condition of selected natural resources. The database provides a summary of, and in many cases allows the user to download a digital copy of, sampling protocols that have been developed by the prototype monitoring parks or other well-established protocols used in national parks. The Protocol Database also makes it possible to download database components (e.g., tables, queries, data entry forms) in MS Access that are consistent with the Natural Resource Database Template that have been developed for a particular protocol. See the Protocol Database website for available protocols (*http://science.nature.nps.gov/im/monitor/protocoldb.cfm*).

NR-GIS Data Store

The NR-GIS Data Store is a key component of the data dissemination strategy employed by the I&M Program. The NR-GIS Data Store is a graphical search interface that links dataset metadata to a searchable data server on which datasets are organized by NPS units, offices and programs. The interface allows customized public or protected searches of natural resource datasets, inventory products and GIS data produced by the I&M and Natural Resource GIS Programs. Each park or network is able to post and curate its data on the server. The NR-GIS Data Store will be integrated with the master NR-GIS Metadata Database application to streamline programmatic data documentation and dissemination processes. The simple browse function of this server can be accessed at: *http://nrdata.nps.gov/*.
See the NR-GIS Data Store website for further information (*http://science.nature.nps.gov/nrdata*).

Appendix D: Project Information Management Process and Work Flow: Deliverables, Repositories, and Guidance Document References

Project phase	Deliverables	Repository	Data management guidance
Planning and approval	• proposal (optional) • study plan, work plan • permit application and permit • contracts and agreements • new record in project management database	• park project funding requests in PMIS/RAMS • permit applications in RPRS • network project management database • working project files, then document archives • final copies of study plan sent to document archives and/or placed in the digital library; records created in NatureBib	• guidelines for data distribution and data discovery • guidelines for organizing project info (including a project file organizer and a template directory structure) • project information management SOP • template language for contracts and agreements (protected information, deliverables) • specifications for deliverables (includes reference to GIS specifications)
Design and testing	• protocol / methodology and SOPs	• documents and supporting materials created and	• naming conventions for

Project phase	Deliverables	Repository	Data management guidance
	• blank field forms and data dictionary • data design documentation • metadata questionnaire – partially complete • Project file documentation – partially complete	maintained in working project folders • final copies of protocols/SOPs sent to document archives and/or placed in the network digital library; records created in NatureBib	files and data objects • data design standards • SOP - metadata procedures and specifications • guidelines for quality assurance and quality control • SOP – GPS data collection and processing
Implementation	• contracts and agreements • raw data forms, field notebooks, trip reports • databases, GIS layers, GPS rover files, list of coordinates • biological specimens, species checklists, catalog of specimen data • photographs, illustrations, sound and video recordings • data certification report	• documents, databases, GPS rover files and GIS layers created and maintained in working project folders • specimens to the park collection or other designated repository	• guidelines for managing third-party data • guidelines for labeling and storage of analytical samples and specimens • standards for voucher collection and cataloguing • photo management strategy • *template language for contracts and agreements* • *naming*

Project phase	Deliverables	Repository	Data management guidance
			conventions for files and data objects • *SOP - metadata procedures and specifications* • *specifications for deliverables (including GIS specs)* • *guidelines for quality assurance and quality control* • *SOP – GPS data collection and processing*
Product delivery and review	• annual report • trend analysis report • final reports - technical or general audiences • publications • completed metadata questionnaire • Project file documentation	• documents and supporting materials created and maintained in working project folders • final copies of reports sent to document archives and/or placed in the network digital library • permit investigator's annual reports to RPRS	• reporting guidelines and templates • *specifications for deliverables (including GIS specs)* • *SOP - metadata procedures and specifications*
Product integration	• finalized metadata posted on the internet • records in searchable corporate databases (e.g., NPSpecies, NatureBib) • products are secure and available	• NPS metadata clearinghouse • NatureBib • NPSpecies • network databases • all remaining materials moved to document archives and stored by calendar year	• integration of project data with corporate data • archival procedures for project records • sensitive info

Project phase	Deliverables	Repository	Data management guidance
			management strategy • NPSpecies and NatureBib guidelines • *project information management SOP* • *guidelines for data distribution and discovery*
Project evaluation and close out	• update record in project management database • completed checklists (data manager, permitting officer, project leader) • documentation of needed modifications	• network project management database • place completed checklist in document archives • maintain modifications documentation with related project documents	• *project information management SOP*

Appendix E: NPS Project Tracking Systems

PMIS

The **Project Management Information System** (**PMIS**) is a Servicewide NPS Intranet application designed to manage information about requests for project funding. It enables parks and NPS offices to submit project proposals to be reviewed, approved and prioritized at park units, regional directorates, and the Washington Office (WASO). PMIS contains historical information for project requests going back to FY99.

In response to a budget call for a particular NPS program for a specific fiscal year (FY), project proposals are submitted, reviewed, approved, prioritized and then formulated under an available funding source by utilizing PMIS. During formulation process for a budget call, a program manager at WASO or a budget officer at a regional directorate determines which project funding requests meet the eligibility criteria for the call to be considered as part of the NPS Budget for a specific FY.

RPRS

The **Research Permit and Reporting System** (RPRS) provides an electronic means to apply for research and collecting permits. Park research coordinators use the system to issue and track all Scientific Research and Collecting Permits. Park research coordinators can use the system to perform the following major actions:

- Receive and organize electronic permit applications, proposals, and peer-reviews from applicants.
- Post and maintain the type of research the park is most interested in attracting.
- Post and maintain park-specific conditions applicable to every permit issued by the park.
- Post an information bulletin used to notify investigators of special conditions or events that could impact planned fieldwork (road closures, area closures, safety-related notices, etc.).
- Process and track permits and denied applications (including revoked or cancelled applications).
- Manage the park Investigator's Annual Report (IAR) database.
- Search the servicewide IAR database.
- Search the servicewide permit database to confirm currently active permits, previously approved studies conducted at other parks, and the reporting of annual accomplishments (submission of IARs) by investigators indicating that they have conducted previous studies in NPS units.

PEPC

The **Planning, Environment, and Public Comment (PEPC)** system is an on-line tool being developed that parks can use to help manage all stages of the compliance process. It includes a "public" side, where anyone can find out about activities going through compliance, and a "private" side for NPS information.

RAMS

The **Resource Activity Management System (RAMS)** is a holistic, activity management system that tracks park resource management from planning stages through the budget process, through work and compliance scheduling, execution and performance to accomplishments. It provides the means to tie resource management activities and their associated actions and components to goals established in broad park planning documents, i.e., General Management Plans (GMP), and to tactical plans, i.e., annual performance plans. It documents what resource activities and actions are planned, what actions are funded, what was actually done, and the results of what was done . RAMS will interface with the Project Management Information System (PMIS) and Operations Formulation System (OFS). It is a joint effort between Cultural Resources and Natural Resources directorates designed for parks to provide full accountability for park resource

management. RAMS is in the final stages of development with a tentative deployment to volunteer pilot parks set for 2004.

Appendix F: Data Stewardship Responsibilities

Role	Programmatic Responsibility	Data Stewardship Responsibilities
Project Crew Member	Collect, record, and verify data	Obtain training in data management for the project. Read and follow project protocols, study plans, and relevant NPS guidance. Communicate with Crew Leader, Project Leader, and Data Manager. Record and verify observed or measured data values. Schedule and perform regular data transfer and backup. Review, verify, and correct field data. Assist with data and procedural documentation, especially deviations from the protocol or study plan.
Project Crew Leader	Supervise crew	Obtain training in data management for the project. Ensure crew members receive data management training and briefings. Read and follow all protocol, project, and relevant Network-level guidelines. Communicate with Crew Members, Project Leader, and Data Manager. Ensure data are regularly transferred, backed up, verified, and entered into the appropriate NPS database(s). Assist with data and procedural documentation.
Data/GIS Specialist or Technician	Process and manage data	Obtain briefings about projects and related data to understand the geospatial and technical requirements and relevance. Communicate with other participants in the project to the extent necessary to accomplish assigned tasks. Perform assigned level of technical data management and/or GIS activities, including data entry, data conversion, and documentation. Work on overall data quality and stewardship with Project Leaders, Resource Specialists, and the Network Data Manager.
Information Technology/ Systems	Provide IT/IS support	Provide and maintain an information systems and technology foundation to support data management.

Role	Programmatic Responsibility	Data Stewardship Responsibilities
Specialist		Advise project participants about capabilities of hardware and software resources to support project and program objectives. Work with Database Manager to resolve hardware and software issues relating to database functions and availability.
Project Leader	Oversee and direct project operations	Ensure Crew Leader receives pertinent training and briefings. Communicate with Crew Leader, Data Manager, and I&M Network Coordinator. Complete project documentation describing the who, what, where, when, why and how of a project. Develop, document and implement standard procedures for field data collection and data handling. Enact and supervise quality assurance and quality control measures for the project. Supervise and certify all field operations, including staff training, equipment calibration, species identification, and data collection. Supervise or perform data entry, verification and validation. Maintain concise explanatory documentation of all deviations from standard procedures. Ensure documentation of important details of each field data collection period. Maintain hard copies of data forms and send original data forms to archive on a regular basis. Work with program coordinators to identify analysis and reporting mechanisms, and to establish a schedule for regular project milestones such as data collection periods, data processing target dates, and reporting deadlines. Produce regular summary reports and conduct periodic trend analysis of data, store the resulting reports, and make them available to users. Act as the main point of contact concerning data content. **The project leader works closely with the data manager to:** Develop quality assurance and quality control procedures specific to project operations.

Role	Programmatic Responsibility	Data Stewardship Responsibilities
		Identify training needs for staff related to data management philosophy, database software use, quality control procedures, etc. Coordinate changes to the field data forms and the user interface for the project database. Fully document and maintain master data. Identify sensitive information that requires special consideration prior to distribution. Manage the archival process to ensure regular archival of project documentation, original field data, databases, reports and summaries, and other products from the project. Define how project data will be transformed from raw data into meaningful information and create data summary procedures to automate and standardize this process. Identify and prioritize legacy data for conversion; convert priority data sets to a modern format. Increase the interpretability and accessibility of existing natural resource information. **Note:** The Project Leader is often a resource specialist, in which case the associated responsibilities for data authority apply (see resource specialist role). A Project Leader without the required background to act as an authority for the data will consult with and involve the appropriate Resource Specialists.
Resource Specialist	Understand the project and make decisions about the data	Understand the objectives of the project, the resulting data, and their scientific and management relevance. Guide development of an Information Needs Assessment based on the objectives of the project. Make decisions about data with regard to validity, utility, sensitivity, and availability. Describe, publish, release, and discuss the data and associated information products. **Note:** The Resource Specialist serving as a Project Leader is also responsible for the duties listed with that role.
GIS Manager	Support park management objectives with GIS and resource	Coordinate and integrate local GIS and resource information management with Network, Regional, and National standards and guidelines. **The GIS specialists will work in collaboration with project**

Role	Programmatic Responsibility	Data Stewardship Responsibilities
	information management	**leaders to:** Determine the GIS data and analysis needs for the project. Develop procedures for field collection of spatial data including the use of GPS and other spatial data collection techniques. Display, analyze, and create maps from spatial data to meet project objectives. Properly document data in compliance with spatial metadata standards. **GIS specialists will also work directly with data managers to:** Design databases and other applications for the network. Create relationships between GIS and non-spatial data and create database and GIS applications to facilitate the integration and analysis of both spatial and non-spatial data. Establish and implement procedures to protect sensitive spatial data according to project needs. Develop and maintain an infrastructure for metadata creation and maintenance. Ensure that project metadata are created and comply with national and agency standards.
Network Data Manager	Ensure inventory and monitoring data are organized, useful, compliant, safe, and available	Assist in developing and implementing procedures to ensure that I&M data collected by NPS staff, cooperators, researchers and others are entered, quality-checked, analyzed, reported, archived, documented, cataloged, and made available to others for management decision-making, research, and education. Provide guidance and support, to the extent possible, to extend Network standards and procedures to studies and data funded by park base and other funding sources to promote integration and availability of datasets. Provide overall Network planning, training, and operational support for the awareness, coordination, integration of data and information management activities, including people, information needs, data, software, and hardware. Serve as Point of Contact for National Park Service database applications (NPSpecies, NatureBib, Dataset Catalog) Coordinate internal and external data management activities.

Role	Programmatic Responsibility	Data Stewardship Responsibilities
		Assign and enforce data stewardship responsibilities. Review and approve all data acquisition plans, hardcopy and electronic field forms, and data dictionaries. Participate in development of Information Needs Assessments. Communicate with Crew Leader, Project Leader, I&M Network Coordinator, and Park GIS/Data Management office. Develop and maintain overall Network and individual Vital Sign data management operating guidelines and relationship to national standards and procedures. Develop and maintain the infrastructure for metadata creation, project documentation, and project data management. Create and maintain project databases in accordance with best practices and current program standards. Provide training in the theory and practice of data management tailored to the needs of project personnel. Develop ways to improve the accessibility and transparency of digital data. Establish and implement procedures to protect sensitive data according to project needs. Collaborate with GIS Specialists to integrate tabular data with geospatial data in a GIS system in a manner that meets project objectives. **Data managers will also work closely with the project leader to:** Define the scope of the project data and create a data structure that meets project needs. Become familiar with how the data are collected, handled, and used. Review quality control and quality assurance aspects of project protocols and standard procedure documentation. Identify elements that can be built into the database structure to facilitate quality control, such as required fields, range limits, pick-lists and conditional validation rules.

Role	Programmatic Responsibility	Data Stewardship Responsibilities
		Create a user interface that streamlines the process of data entry, review, validation, and summarization that is consistent with the capabilities of the project staff. Develop automated database procedures to improve the efficiency of the data summarization and reporting process. Make sure that project documentation is complete, complies with metadata requirements, and enhances the interpretability and longevity of the project data. Ensure regular archival of project materials. Inform project staff of changes and advances in data management practices. Additional examples of the duties and responsibilities of the network data managers are listed in I&M Program Vision and Organizational Framework document "Network Data Manager Overview of Responsibilities". **NOTE:** Data Managers with Prototype Monitoring Programs have the same basic duties and responsibilities as the network data managers but also are responsible for mentoring and training others and developing and testing new approaches to data analysis, synthesis, and reporting of monitoring results.
Database Manager	Know and use databases and applications	Install, maintain, and support specific database software applications and NPS database applications. Work with Information Technology Specialists to resolve hardware and software issues.
Curator	Oversee all aspects of the acquisition, documentation, preservation, and use of park collections	Know park natural resource collections Conduct accessioning, cataloging, legal, and other documentation of collections Manage collections databases Recognize objects needing conservation treatment Recommend and refer treatment to the appropriate facility Work with Network Data Manager to acquire and process data related to natural resource collections
Statistician or Biometrician	Analyze data and present information	Work with the Network Ecologist to analyze and report data according to established protocols. Work with the Network Data Manager to acquire and process raw data from databases and store derived data and

Role	Programmatic Responsibility	Data Stewardship Responsibilities
		information after analysis
Network Ecologist	Integrate science in network activities	Ensure useful data are collected and managed by integrating natural resource science in network activities and products, including objective setting, sample design, data analysis, synthesis, and reporting. Assist with development and modification of monitoring protocols and inventory study plans. Work with the Network Data Manager to incorporate data management in monitoring protocols. Participate in the development of Information Needs Assessments based on the objectives of the project. Guide and/or perform statistical and other analyses of network data. Contribute to the synthesis and reporting of data and information. Provide guidance and support, to the extent possible, to extend Network standards and procedures to studies and data funded by park base and other funding sources to promote integration and availability of datasets.
Network Coordinator	Coordinate all network activities	Ensure programmatic data and information management requirements are met as part of overall Network business. Communicate with Network staff, park staff at all levels, and other appropriate audiences to support and emphasize data management as a critical aspect of network business Work with Network Data Manager regarding data management policy and guidelines, budget, staffing, and training. Hold Network staff accountable for responsibilities involving data management.
I&M Data Manager (National Level)	Provide service-wide database availability and support	Provide services to receive, convert, store, and archive data in service-wide databases. Work with Network Data Manager to resolve local issues involving the access and use of inventory and monitoring databases. Provide training where possible. Design and maintain standardized, master databases for Servicewide planning, decision-making, and accountability (e.g.,

Role	Programmatic Responsibility	Data Stewardship Responsibilities
		NPSpecies, NatureBib, Dataset Catalog, Database Template, GIS tools). Collaborate with networks to help develop overall data management vision and approach, and continual improvement of specific tools. Coordinate establishment of standards for naming conventions and content of data management plans and monitoring protocols. Promote collaboration and integration with other divisions and programs including the GIS community, fire program, air resources, water resources, geologic resources, etc. Facilitate coordination and collaboration among the parks and networks by providing examples of good database designs with flexibility to allow adjustments for different situations.
Other End Users	Use and apply Network services and products	These 'information consumers' include park managers and superintendents, researchers, staff from other agencies, and the public. End users at all levels are generally responsible for providing necessary and requested feedback, review, and comments on various products in order to sustain the continuous improvement of network operations and services. End users are responsible for the appropriate use and application of data and derived products.

Appendix G: I&M Program Network Data Manager Position Description

OVERVIEW	***Foster the success of the I&M program through data stewardship; ensure quality data sets throughout the network, integrate data and coordinate and communicate openly with all data managers throughout the region, and contribute where appropriate to the national I&M efforts.***

	LEVEL OF RESPONSIBILITY	
DUTY	*RESPONSIBLE FOR* ***(core duty)***	*INVOLVED IN/AWARE OF* ***(secondary duty)***
Design Databases	• Understand and follow conventions of the I&M Database Template • Ensure "good" relational database modeling practices are followed for all I&M projects across the network • Ensure synchronization of tabular and spatial data models for all I&M projects across the network • Coordinate efforts between similar data models within the network • Ensure that appropriate design documentation has been created, such as data dictionaries and model diagrams • Ensure database designs are compatible with field collection needs	• Assist with non-I&M projects to ensure quality database designs • Coordinate efforts between similar data models across all networks in region • Contribute to national efforts of designing cohesive datasets

DUTY	LEVEL OF RESPONSIBILITY	
	RESPONSIBLE FOR ***(core duty)***	*INVOLVED IN/AWARE OF* ***(secondary duty)***
Implement Databases / Maintain Datasets	• Ensure appropriate database software is selected for all I&M projects across network • Ensure suitable cataloging, storage, and accessibility of digital data • Develop archiving procedures and ensure their practice • Evaluate legacy data for usability • Develop and implement QA/QC procedures and data validation tools	• Coordinate efforts between similar datasets across all networks in the region
Populate Data Sets	• Ensure reliability of "location" data collection methods for accurate spatial data • Ensure data integrity as data sets are converted to electronic format • Provide programming skills for data entry forms and other input methods • Ensure FDGC-compliant metadata for spatial datasets • Generate spatial datasets from tabular data • Analyze data to identify potential data anomalies • Ensure population of national datasets such as Dataset Catalog, NPSpecies, ANCS+, and NPBib • Ensure data allows for appropriate setup of ArcView to Access Links	• Contribute to national efforts of collecting cohesive datasets

DUTY	***LEVEL OF RESPONSIBILITY***	
	RESPONSIBLE FOR ***(core duty)***	*INVOLVED IN/AWARE OF* ***(secondary duty)***
Provide Data Sets/Products	• Assist others in the understanding and analysis of collected data • Format data sets as required by others for reporting or analysis tools • Develop desktop (distributable) and web-based applications for data entry, viewing, and reporting • Develop GIS products for data analysis and dissemination	• Assist statisticians with developing data formats to fit statistical models or analysis packages • Assist with creation of ArcView to Access links • Assist with creation of graphics presentations

NON-DUTIES	***LEVEL OF RESPONSIBILITY***
	NOT RESPONSIBLE FOR
	• Park GIS and database solutions outside of the I&M Program; should assist when skills and expertise are required, but should not be solely responsible for implementing projects at the park level • System or network administration and support, including hardware and software issues • General data entry; should not be responsible for keying data in from field forms • Metadata and documentation generation

Appendix H. Metadata Generation Tools

Dataset Catalog: Dataset Catalog is a tool for cataloging abbreviated metadata on geospatial and biological data sets pertaining to park(s) and/or a network. It provides parks and/or networks a means whereby they can inventory, organize, and maintain information about data set holdings locally. While Dataset Catalog is not intended to be an exhaustive metadata listing, it does assist parks and networks in beginning to meet the mandates of EO 12906. With the current version of Dataset Catalog (version 2), records can be exported as an FGDC text file, which can than be imported into other metadata tools. Version 2.1 (in development) will include the ability to export records in Extensible Markup Language (XML). The I&M Program recommends that all relevant datasets at I&M parks and networks be cataloged in at least simple Dataset Catalog format.

Spatial Metadata Management System: SMMS is a tool with the capability to create, edit, view, and publish metadata that is compliant with FGDC requirements. SMMS uses an MS Access database structure combined with an advanced FGDC-compliant metadata editor. The software allows selection of views depending on whether the user wants the full standard, biological, or the minimal compliant view of Sections 1 and 7. There is online Help to describe the purpose, usage or mandatory status of metadata elements. The context-sensitive help file provides the FGDC definition for each field on the screen. In addition to Help files, there are sample metadata records for most sections that provide "real world" examples. The NPS Integrated Metadata System Plan recommends SMMS for FGDC Biological Profile and other geospatial metadata creation. The CAKN Prototype GIS staff have been using SMMS for several years. However, this platform may be dropped in the near future as customization of ArcCatalog metadata tools make it the optimal tool.

ArcCatalog: ArcCatalog is a management tool for GIS files contained within the ArcGIS Desktop suite of applications. With ArcCatalog, users can browse, manage, create, and organize tabular and GIS data. In addition, ArcCatalog comes with support for several popular metadata standards that allow one to create, edit, and view information about the data. There are editors to enter metadata, a storage schema, and property sheets to view the data. With ArcCatalog users can view GIS data holdings, preview geographic information, view and edit metadata, work with tables, and define the schema structure for GIS data layers. Metadata within ArcCatalog is stored exclusively as Extensible Markup Language (XML) files. The NPS Integrated Metadata System Plan recommends ArcCatalog for gathering GIS-integrated geospatial metadata. An optional but highly recommended extension for ArcCatalog is the NPS Metadata ArcCatalog Extension developed by NPS Midwest Region GIS Technical Support Center. The extension fixes several ArcGIS 8 metadata errors and provides added functionality for NPS users. Development is also underway to provide Biological Profile editing capability and NPS Profile support. This tool is under review by CAKN Prototype GIS staff.

Metadata Parser: The MetaParser (mp) program is used to validate metadata records by checking the syntax against the CSDGM and to generate compliant output files for posting to clearinghouses. It generates a textual report indicating errors in the metadata,

primarily in the structure, but also in the values of some of the scalar elements where values are restricted by the standard.

Appendix I. Project Archive Checklist

National Park Service
Inventory & Monitoring Program
Central Alaska Network

Project Check List

Place a copy of this completed form in the hardcopy folder of the project. Store an electronic copy of this document with appropriate computer files.

Project Name: _____
I&M Project Number: _____

Project Personnel: _____

Project Contact (email/phone):_____

Date (form completed): _____

Location of Materials:
Physical files: _____
Electronic files: _____

Check List:

√	Project Elements	File Type*	Entered Redisc.	NPS Archive†	Comments/Specific Location (of either physical or digital items):
	Final Proposal				
	Study Plan / Protocols				
	Permit				
	Compliance Docs				
	Field Forms†				
	Field Notes†				
	Tabular Data (databases, spreadsheets, etc.)				File name and location:

	Reports				
	Physical Maps†				
	GIS Files				
	Photographs				
	Other Files (presentations, posters, graphics, related literature, etc.)				
	Digital backup of project†				(circle as appropriate): DVD CD Other:
	HD backup (shared drive)				
	Reports to Park Library and ARLIS				
	NatureBib Entry				
	NPSpecies Entry				
	Metadata (digital and hardcopy)				
	Closeout of Project/Permits				

*File Type: PF = Project Folder (hardcopy in project folder), RF = Park Resource Library Folder (hardcopy in library folder), PE = Project Electronic Folder, O = Other (specify)

Notes:

Appendix J: Glossary

Data **–** Observational information gathered directly (in the field) from a natural resource via specific protocols and organized for analysis, summary or reporting.

Data may be in either analog or digital form (generally stored either on paper or a variety of computer-compatible media), though the latter is encouraged where ever feasible. Data may exist in several states (conditions) including 'raw', 'validated' and 'analyzed'. 'Analyzed' data includes 'reported' or 'summarized' data and may represent 'information' as a final form of the data from which decisions or conclusions may be made. Ultimately, data are intended to contribute to the knowledge and decisions regarding the conditions, processes, and changes within the ecosystem.

Dataset **–** A *dataset* can best be considered a convenient grouping of data, or individual observations, such that the summary of the information will be meaningful to prospective users.

Information **–** Material that is either derived from data or supports in a meaningful way the interpretation and/or understanding of data. *Information* may take the form of reports that draw conclusions based on data, database documentation, metadata, users guides for data analysis/browsing/dissemination applications, project timelines/study plans/etc.

Raw Data **–** Data obtained from the environment and that has not been subjected to any quality assurance or control beyond those applied during field work. Typically, raw data constitutes field data sheets but may also include remotely sensed data, data gathered electronically on field computers and photographic imagery.

Validated Data **–** Data that have been verified according to the standard operating procedure under which the data were gathered (typically the protocol for a given project) and are deemed ready for reporting and/or analysis.

Analyzed Data **– Dat**a that have been subjected to analytical routines after field collection and verification. This includes laboratory results as well as statistical operations conducted on the data for the purposes of arriving at a measure of the given ecological parameter.

Legacy Data **– Data collected before the CAKN Data Management Plan was implemented. Generally, this refers to data collected in the park units that the CAKN deems important to the monitoring mission.**

Program Output **–** The combination of data and information generated or otherwise maintained by I&M Program activities.

Data Cleansing **–** The process of initial data quality assurance and control performed immediately after data entry. Examples of data cleansing include removing typographical errors made on field sheets and during data entry and removing spurious characters from automated data routines such as satellite download.

NatureBib **–** The National Park Service bibliographic database. NatureBib is designed to work with NPSFocus to help staff learn of relevant reports, locate them and where possible obtain digitally.

NR-GIS **–** The National Park Service natural resource and GIS data store. The data store is designed primarily as a metadata browser for data discovery but also serves as a data source for select tabular and spatial data.

***NPSFocus* –** The National Park Service online library designed to provide a medium for literature research and, where possible, digital access to relevant reports.